```
GAATTCGGAG GAATTATTCA AAACATAAAC ACAATAAACA ATTTGAGTAG TTGCCGCACA    60

CACACACACA CACAGCCCGT GGATTATTAC ACTAAAAGCG ACACTCAATC CAAAAAATCA   120

GCAACAAAAA CATCAATAAA C ATG CAT TGG ATT AAA TGT TTA TTA ACA GCA    171
                        Met His Trp Ile Lys Cys Leu Leu Thr Ala
                         1                5                    10

TTC ATT TGC TTC ACA GTC ATC GTG CAG GTT CAC AGT TCC GGC AGC TTT    219
Phe Ile Cys Phe Thr Val Ile Val Gln Val His Ser Ser Gly Ser Phe
             15                20                25

GAG TTG CGC CTG AAG TAC TTC AGC AAC GAT CAC GGG CGG GAC AAC GAG    267
Glu Leu Arg Leu Lys Tyr Phe Ser Asn Asp His Gly Arg Asp Asn Glu
         30                35                40

GGT CGC TGC TGC AGC GGG GAG TCG GAC GGA GCG ACG GGC AAG TGC CTG    315
Gly Arg Cys Cys Ser Gly Glu Ser Asp Gly Ala Thr Gly Lys Cys Leu
         45                50                55

GGC AGC TGC AAG ACG CGG TTT CGC GTC TGC CTA AAG CAC TAC CAG GCC    363
Gly Ser Cys Lys Thr Arg Phe Arg Val Cys Leu Lys His Tyr Gln Ala
     60                65                70

ACC ATC GAC ACC ACC TCC CAG TGC ACC TAC GGG GAC GTG ATC ACG CCC    411
Thr Ile Asp Thr Thr Ser Gln Cys Thr Tyr Gly Asp Val Ile Thr Pro
75                80                85                90

ATT CTC GGC GAG AAC TCG GTC AAT CTG ACC GAC GCC CAG CGC TTC CAG    459
Ile Leu Gly Glu Asn Ser Val Asn Leu Thr Asp Ala Gln Arg Phe Gln
             95                100                105

AAC AAG GGC TTC ACG AAT CCC ATC CAG TTC CCC TTC TCG TTC TCA TGG    507
Asn Lys Gly Phe Thr Asn Pro Ile Gln Phe Pro Phe Ser Phe Ser Trp
         110                115                120
```

FIG.1A

```
CCG GGT ACC TTC TCG CTG ATC GTC GAG GCC TGG CAT GAT ACG AAC AAT    555
Pro Gly Thr Phe Ser Leu Ile Val Glu Ala Trp His Asp Thr Asn Asn
        125                 130                 135

AGC GGC AAT GCG CGA ACC AAC AAG CTC CTC ATC CAG CGA CTC TTG GTG    603
Ser Gly Asn Ala Arg Thr Asn Lys Leu Leu Ile Gln Arg Leu Leu Val
        140                 145                 150

CAG CAG GTA CTG GAG GTG TCC TCC GAA TGG AAG ACG AAC AAG TCG GAA    651
Gln Gln Val Leu Glu Val Ser Ser Glu Trp Lys Thr Asn Lys Ser Glu
155                 160                 165                 170

TCG CAG TAC ACG TCG CTG GAG TAC GAT TTC CGT GTC ACC TGC GAT CTC    699
Ser Gln Tyr Thr Ser Leu Glu Tyr Asp Phe Arg Val Thr Cys Asp Leu
                175                 180                 185

AAC TAC TAC GGA TCC GGC TGT GCC AAG TTC TGC CGG CCC CGC GAC GAT    747
Asn Tyr Tyr Gly Ser Gly Cys Ala Lys Phe Cys Arg Pro Arg Asp Asp
                190                 195                 200

TCA TTT GGA CAC TCG ACT TGC TCG GAG ACG GGC GAA ATT ATC TGT TTG    795
Ser Phe Gly His Ser Thr Cys Ser Glu Thr Gly Glu Ile Ile Cys Leu
                205                 210                 215

ACC GGA TGG CAG GGC GAT TAC TGT CAC ATA CCC AAA TGC GCC AAA GGC    843
Thr Gly Trp Gln Gly Asp Tyr Cys His Ile Pro Lys Cys Ala Lys Gly
        220                 225                 230

TGT GAA CAT GGA CAT TGC GAC AAA CCC AAT CAA TGC GTT TGC CAA CTG    891
Cys Glu His Gly His Cys Asp Lys Pro Asn Gln Cys Val Cys Gln Leu
235                 240                 245                 250

GGC TGG AAG GGA GCC TTG TGC AAC GAG TGC GTT CTG GAA CCG AAC TGC    939
Gly Trp Lys Gly Ala Leu Cys Asn Glu Cys Val Leu Glu Pro Asn Cys
                255                 260                 265
```

FIG.1B

```
ATC CAT GGC ACC TGC AAC AAA CCC TGG ACT TGC ATC TGC AAC GAG GGT      987
Ile His Gly Thr Cys Asn Lys Pro Trp Thr Cys Ile Cys Asn Glu Gly
        270                 275                 280

TGG GGA GGC TTG TAC TGC AAC CAG GAT CTG AAC TAC TGC ACC AAC CAC     1035
Trp Gly Gly Leu Tyr Cys Asn Gln Asp Leu Asn Tyr Cys Thr Asn His
        285                 290                 295

AGA CCC TGC AAG AAT GGC GGA ACC TGC TTC AAC ACC GGC GAG GGA TTG     1083
Arg Pro Cys Lys Asn Gly Gly Thr Cys Phe Asn Thr Gly Glu Gly Leu
    300                 305                 310

TAC ACA TGC AAA TGC GCT CCA GGA TAC AGT GGT GAT GAT TGC GAA AAT     1131
Tyr Thr Cys Lys Cys Ala Pro Gly Tyr Ser Gly Asp Asp Cys Glu Asn
315                 320                 325                 330

GAG ATC TAC TCC TGC GAT GCC GAT GTC AAT CCC TGC CAG AAT GGT GGT     1179
Glu Ile Tyr Ser Cys Asp Ala Asp Val Asn Pro Cys Gln Asn Gly Gly
            335                 340                 345

ACC TGC ATC GAT GAG CCG CAC ACA AAA ACC GGC TAC AAG TGT CAT TGC     1227
Thr Cys Ile Asp Glu Pro His Thr Lys Thr Gly Tyr Lys Cys His Cys
            350                 355                 360

GCC AAC GGC TGG AGC GGA AAG ATG TGC GAG GAG AAA GTG CTC ACG TGT     1275
Ala Asn Gly Trp Ser Gly Lys Met Cys Glu Glu Lys Val Leu Thr Cys
            365                 370                 375

TCG GAC AAA CCC TGT CAT CAG GGA ATC TGC CGC AAC GTT CGT CCT GGC     1323
Ser Asp Lys Pro Cys His Gln Gly Ile Cys Arg Asn Val Arg Pro Gly
        380                 385                 390

TTG GGA AGC AAG GGT CAG GGC TAC CAG TGC GAA TGT CCC ATT GGC TAC     1371
Leu Gly Ser Lys Gly Gln Gly Tyr Gln Cys Glu Cys Pro Ile Gly Tyr
395                 400                 405                 410
```

FIG.1C

3

```
AGC GGA CCC AAC TGC GAT CTC CAG CTG GAC AAC TGC AGT CCG AAT CCA       1419
Ser Gly Pro Asn Cys Asp Leu Gln Leu Asp Asn Cys Ser Pro Asn Pro
            415                   420                   425

TGC ATA AAC GGT GGA AGC TGT CAG CCG AGC GGA AAG TGT ATT TGC CCA       1467
Cys Ile Asn Gly Gly Ser Cys Gln Pro Ser Gly Lys Cys Ile Cys Pro
            430                   435                   440

GCG GGA TTT TCG GGA ACG AGA TGC GAG ACC AAC ATT GAC GAT TGT CTT       1515
Ala Gly Phe Ser Gly Thr Arg Cys Glu Thr Asn Ile Asp Asp Cys Leu
            445                   450                   455

GGC CAC CAG TGC GAG AAC GGA GGC ACC TGC ATA GAT ATG GTC AAC CAA       1563
Gly His Gln Cys Glu Asn Gly Gly Thr Cys Ile Asp Met Val Asn Gln
            460                   465                   470

TAT CGC TGC CAA TGC GTT CCC GGT TTC CAT GGC ACC CAC TGT AGT AGC       1611
Tyr Arg Cys Gln Cys Val Pro Gly Phe His Gly Thr His Cys Ser Ser
475                         480         485                   490

AAA GTT GAC TTG TGC CTC ATC AGA CCG TGT GCC AAT GGA GGA ACC TGC       1659
Lys Val Asp Leu Cys Leu Ile Arg Pro Cys Ala Asn Gly Gly Thr Cys
                  495               500                   505

TTG AAT CTC AAC AAC GAT TAC CAG TGC ACC TGT CGT GCG GGA TTT ACT       1707
Leu Asn Leu Asn Asn Asp Tyr Gln Cys Thr Cys Arg Ala Gly Phe Thr
                  510               515                   520

GGC AAG GAT TGC TCT GTG GAC ATC GAT GAG TGC AGC AGT GGA CCC TGT       1755
Gly Lys Asp Cys Ser Val Asp Ile Asp Glu Cys Ser Ser Gly Pro Cys
            525                   530                   535

CAT AAC GGC GGC ACT TGC ATG AAC CGC GTC AAT TCG TTC GAA TGC GTG       1803
His Asn Gly Gly Thr Cys Met Asn Arg Val Asn Ser Phe Glu Cys Val
            540                   545                   550
```

FIG.1D

```
TGT GCC AAT GGT TTC AGG GGC AAG CAG TGC GAT GAG GAG TCC TAC GAT        1851
Cys Ala Asn Gly Phe Arg Gly Lys Gln Cys Asp Glu Glu Ser Tyr Asp
555             560                 565                 570


TCG GTG ACC TTC GAT GCC CAC CAA TAT GGA GCG ACC ACA CAA GCG AGA        1899
Ser Val Thr Phe Asp Ala His Gln Tyr Gly Ala Thr Thr Gln Ala Arg
                575                 580                 585


GCC GAT GGT TTG ACC AAT GCC CAG GTA GTC CTA ATT GCT GTT TTC TCC        1947
Ala Asp Gly Leu Thr Asn Ala Gln Val Val Leu Ile Ala Val Phe Ser
                590                 595                 600


GTT GCG ATG CCT TTG GTG GCG GTT ATT GCG GCG TGC GTG GTC TTC TGC        1995
Val Ala Met Pro Leu Val Ala Val Ile Ala Ala Cys Val Val Phe Cys
                605                 610                 615


ATG AAG CGC AAG CGT AAG CGT GCT CAG GAA AAG GAC GAC GCG GAG GCC        2043
Met Lys Arg Lys Arg Lys Arg Ala Gln Glu Lys Asp Asp Ala Glu Ala
                620                 625                 630


AGG AAG CAG AAC GAA CAG AAT GCG GTG GCC ACA ATG CAT CAC AAT GGC        2091
Arg Lys Gln Asn Glu Gln Asn Ala Val Ala Thr Met His His Asn Gly
635                 640                 645                 650


AGT GGG GTG GGT GTA GCT TTG GCT TCA GCC TCT CTG GGC GGC AAA ACT        2139
Ser Gly Val Gly Val Ala Leu Ala Ser Ala Ser Leu Gly Gly Lys Thr
                655                 660                 665


GGC AGC AAC AGC GGT CTC ACC TTC GAT GGC GGC AAC CCG AAT ATC ATC        2187
Gly Ser Asn Ser Gly Leu Thr Phe Asp Gly Gly Asn Pro Asn Ile Ile
                670                 675                 680


AAA AAC ACC TGG GAC AAG TCG GTC AAC AAC ATT TGT GCC TCA GCA GCA        2235
Lys Asn Thr Trp Asp Lys Ser Val Asn Asn Ile Cys Ala Ser Ala Ala
                685                 690                 695
```

FIG.1E

```
GCA GCG GCG GCG GCG GCA GCA GCG GCG GAC GAG TGT CTC ATG TAC GGC    2283
Ala Ala Ala Ala Ala Ala Ala Ala Ala Asp Glu Cys Leu Met Tyr Gly
    700                 705                 710

GGA TAT GTG GCC TCG GTG GCG GAT AAC AAC AAT GCC AAC TCA GAC TTT    2331
Gly Tyr Val Ala Ser Val Ala Asp Asn Asn Asn Ala Asn Ser Asp Phe
715                 720                 725                 730

TGT GTG GCT CCG CTA CAA AGA GCC AAG TCG CAA AAG CAA CTC AAC ACC    2379
Cys Val Ala Pro Leu Gln Arg Ala Lys Ser Gln Lys Gln Leu Asn Thr
                735                 740                 745

GAT CCC ACG CTC ATG CAC CGC GGT TCG CCG GCA GGC AGC TCA GCC AAG    2427
Asp Pro Thr Leu Met His Arg Gly Ser Pro Ala Gly Ser Ser Ala Lys
                750                 755                 760

GGA GCG TCT GGC GGA GGA CCG GGA GCG GCG GAG GGC AAG AGG ATC TCT    2475
Gly Ala Ser Gly Gly Gly Pro Gly Ala Ala Glu Gly Lys Arg Ile Ser
                765                 770                 775

GTT TTA GGC GAG GGT TCC TAC TGT AGC CAG CGT TGG CCC TCG TTG GCG    2523
Val Leu Gly Glu Gly Ser Tyr Cys Ser Gln Arg Trp Pro Ser Leu Ala
    780                 785                 790

GCG GCG GGA GTG GCC GGA GCC TGT TCA TCC CAG CTA ATG GCT GCA GCT    2571
Ala Ala Gly Val Ala Gly Ala Cys Ser Ser Gln Leu Met Ala Ala Ala
795                 800                 805                 810

TCG GCA GCG GGC AGC GGA GCG GGG ACG GCG CAA CAG CAG CGA TCC GTG    2619
Ser Ala Ala Gly Ser Gly Ala Gly Thr Ala Gln Gln Gln Arg Ser Val
                815                 820                 825

GTC TGC GGC ACT CCG CAT ATG TAACTCCAAA AATCCGGAAG GGCTCCTGGT       2670
Val Cys Gly Thr Pro His Met
                830
AAATCCGGAG AAATCCGCAT GGAGGAGCTG ACAGCACATA CACAAAGAAA AGACTGGGTT  2730
GGGTTCAAAA TGTGAGAGAG ACGCCAAAAT GTTGTTGTTG ATTGAAGCAG TTTAGTCGTC  2790
ACGAAAAATG AAAAATCTGT AACAGGCATA ACTCGTAAAC TCCCTAAAAA ATTTGTATAG  2850
TAATTAGCAA AGCTGTGACC CAGCCGTTTC GATCCCGAAT TC                     2892
```

FIG.1F

6

TM

SP   EGF   N   cdc10 PA opa

Construct	% AGGREGATION WITH DI	WITH Ser
1. pMtNMg	40	21
2. △Sph	0	nt
3. △Cla	0	nt
4. △EGF(7–17)	0	nt
5. △EGF(9–26)	0	nt
6. △EGF(17–30)	22	nt
7. △EGF(7–9)	20	14
8. △EGF(9–17)	0	0
9. △EGF(17–26)	10	8
10. △EGF(26–30)	5	7
11. △EGF(9–30)	0	nt
12. △EGF(7–26)	0	nt
13. △Cla+EGF(9–17)	35	20
14. △Cla+EGF(17–26)	0	nt
15. SPLIT	42	nt
16. △Cla+EGF(9–13)	47	25

FIG.2A

7

FIG.2B

17. ΔCla+EGF(11–15)	0	0
18. ΔCla+EGF(13–17)	0	nt
19. ΔCla+EGF(10–13)	56	23
20. ΔCla+EGF(11–13)	0	nt
21. ΔCla+EGF(10–12)	0	nt
22. ΔCla+EGF(10–11)	0	nt
23. ΔCla+EGF(10–12)	45	nt
24. ΔCla+EGF(11–12)	11	nt
25. ΔEGF	0	nt
26. ΔEGF+EGF(9–17)	24	nt
27. ΔEGF+EGF(9–13)	40	nt
28. ΔEGF+EGF(10–13)	45	23
29. ΔEGF+EGF(10–12)	48	nt
30. ΔECN	0	nt
31. ΔECN+EGF(10–13)	26	nt
32. ΔECN+EGF(10–12)	47	22
33. ΔCla+XEGF(10–13)	42	20

8

FIG.3

DROSOPHILA ── EGF11 ──

** ** ** ** ** ** ** ** ** ** ** ** **

EDIDECDQGS PCEHNGICVNTPGSYRCNCSQGFTGPRCETNINECESHPCQNEGSCLDDPGTFRCVCMPGFTGTQCE

XENOPUS

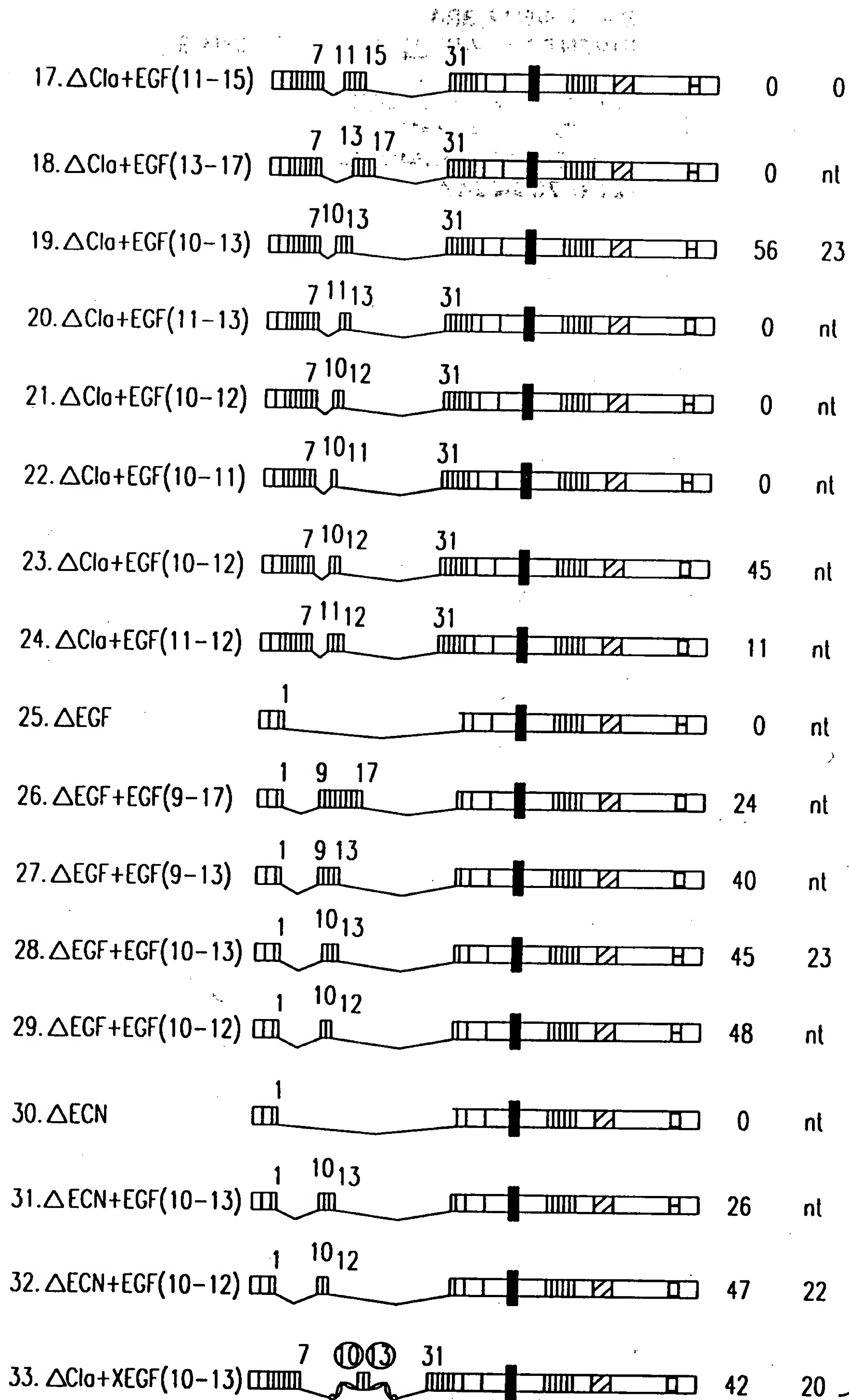NDVDECSLGANPCEHGGRCTNTLGSFQCNCPQGYAGPRCEIDVNECLSNPCQNDSTCLDQIGEFQCICMPGYEGLYCE
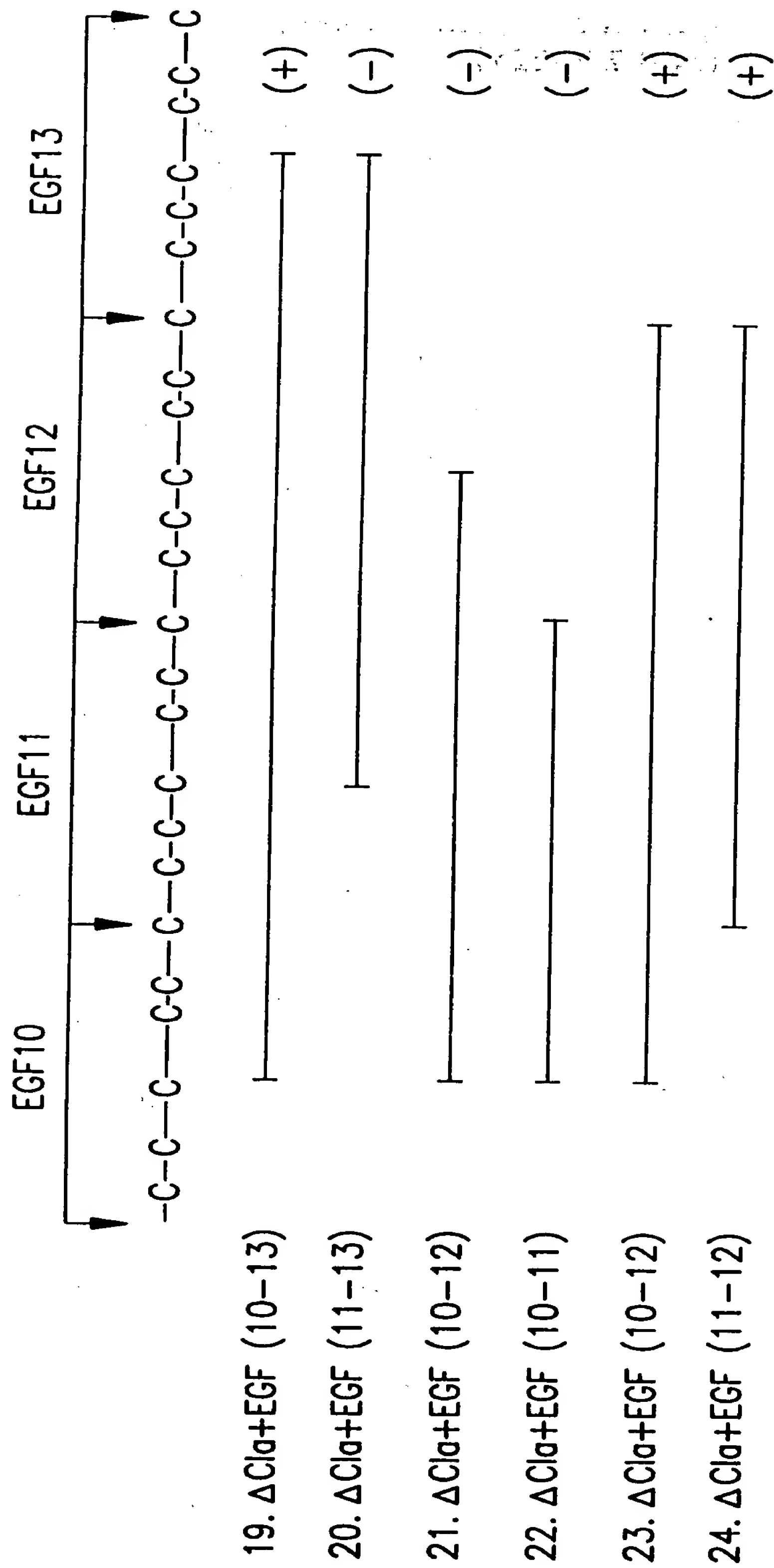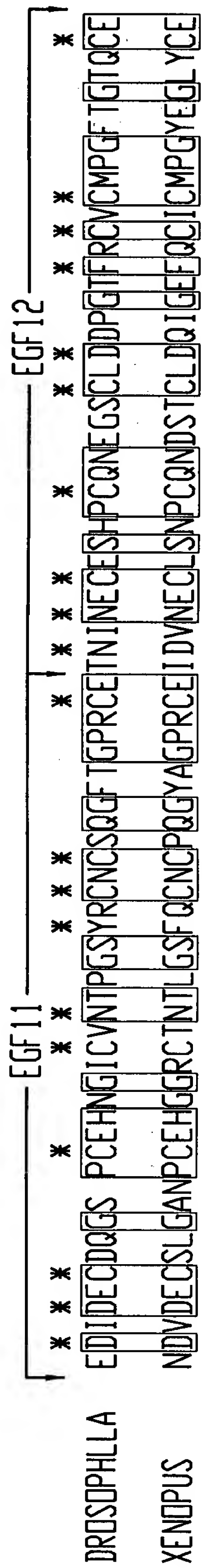
── EGF12 ──

P

FIG.4

```
1    CCGAGTCGAGCGCCGTGCTTCGAGGGGTGATGAGCCCTTTCTGTCAACGCTAAAGATC
121  AAGCACATACTAAGGTCCATATAAATAATAATAATTGTGTGATAACAACATTAT
241  GGCCGTTATTCAGCTATCCAGAGCAAGTGTAGTGTGGCAAAATAGAAACAAAGGCA
361  CAATCCAGAGTGAATCCGAAACAAACTCCATCTAGATCGCCAACCAGCATCACGCTCGCA

481  TCGTCGTTGGAGTCAACAATAGAATCAGCAGACAGCCTGGGAATGTCCAAGAAGACGGCG
     SerSerLeuGluSerThrIleGluSerAlaAspSerLeuGlyMetSerLysLysThrAla

601  CGGGATTGTCGATCATTAAAGTCTGCCTGCAACTTAATTGCTTTAATTTAATACTGTTA
     ArgAspCysArgSerLeuLysSerAlaCysAsnLeuIleAlaLeuIleLeuLeuLeu
     - - - - - - - - - - - - - - - - - - - - - - - - -

721  AACAGGCCATCTACTCAAGGGCTATTGCTGCGGCATGCCAGCGGAACTTAGGCCACCAAG
     AsnSerHisLeuLeuAsnGlyTyrCysCysGlyMetProAlaGluLeuArgAlaThrLys
                                                              #2
841  ACGGAGCAGGTGCCAGCATATCCACGGGCTGTTCGTTTGGCAAGGCCACCACCAAGATA
     ThrGluGlnGlyAlaSerIleSerThrGlyCysSerPheGlyAsnAlaThrThrLysIle

961  ACGTTTCGTTGGACGAAGTCGTTTACGCTGGATACTGCAGGCGTTGGATATGTACAACACA
     ThrPheArgTrpThrLysSerPheThrLeuIleLeuGlnAlaLeuAspMetTyrAsnThr
     #3
1081 TCGCCCGGAGTGGAAGACGCTGGACCACATCGGGGCGGGATCACCTACCGTGTC
     SerProGluTrpLysThrLeuAspHisIleGlyArgAsnAlaArgIleThrTyrArgVal

1201 GACGGATCAGTTCGGTCACTACGCCTGGCGCGTACGGGTCAGAAGCTCTGCCTGAATGGC
     AspAspGlnPheGlyHisTyrAlaCysGlySerGluGlyGlnLysLeuCysLeuAsnGly
```

FIG.5A

TACAAAAACATCAGCGCCTATCAAGTGGAAGTGTCAAGTGTGAACAAAAACAAAAACGAGAG
CCAAACAAACCAAACAAACGAGAGGCAAAGTGGAGAAAATGATACAGCATCCAGAGTAC
CCAAAATCTGCATACATGGGCTAATTAAGGCTGCCCAGGAATTACATTGTGTGGTGC
AACGCCCCCAGAATGTACAAAAATGTTTAGGAAACATTTCGGCGAAAACCAGTACGTCG
                                    MetPheArgLysHisPheArgArgLysProAlaThrSer 13

ACAAAAAGGCAGGTCCGAGGCATCGGGTACCCAAAATGCGACCCTGCCATCGACGATC
ThrLysArgGlnArgProArgHisArgValProLysIleAlaThrLeuProSerThrIle 53

GTCCATAAGATATCCGGCAGCTGGTAACTTCGAGCTGGAAATATTAGAAATCTCAAATACC
ValHisLysIleSerAlaAlaAlaGlyAsnPheGluLeuGluIleLeuGluIleSerAsnThr 93
                                                              #1

ACGATAGGCTGCTGCCATGCACGACGGCATTCCGGCTGTGCCTGAAGGAGTACCAGACC
ThrIleGlyCysSerProCysThrThrAlaPheArgLeuCysLeuLysGluTyrGlnThr 133

CTGGGTGGCTCCAGCTTTGTGCTCAGGATCCGGGTGTGGGAGCCATTGTGCTGCCCTTT
LeuGlyGlySerSerPheValLeuSerAspProGlyValGlyAlaIleValLeuProPhe 173

TCCTATCCAGATGCGGAGAGGTTAATTGAGGAAACATCATACTCGGGGGTGATACTGCCG
SerTyrProAspAlaGluArgLeuIleGluGluThrSerTyrSerGlyValIleLeuPro 213

CGGGTGCAATGCGCCGTTACCTACTACAACACGACCTGCACGACCTTGCCGTCGCGGG
ArgValGlnCysAlaValThrTyrTyrAsnThrThrCysThrThrPheCysArgProArg 253
                                                           #4

TGGCAGGGCGTCAACTGCGAGGAGGCCATATGCAAGGCGGGCTGCGACCCGTCCACGGC
TrpGlnGlyValAsnCysGluGluAlaIleCysLysLysAlaGlyCysAspProValHisGly 293

FIG.5B

12

FIG.6

5' END OF CODING REGION

3' END OF CODING REGION

400bp

200bp

1 2 3

h NOTCH (~10k)

hN4k

hN3k

hN2k

hN5k

FIG.7

HASH VALUE= 1
PROCESSING= ALIGN

hN3k= 3 5'-3' COMPOSITE
hN2k= 2 5'-3' COMPOSITE
hN5k= 5 5'-3' COMPOSITE
hn4k= 4 n'-x' COMPOSITE
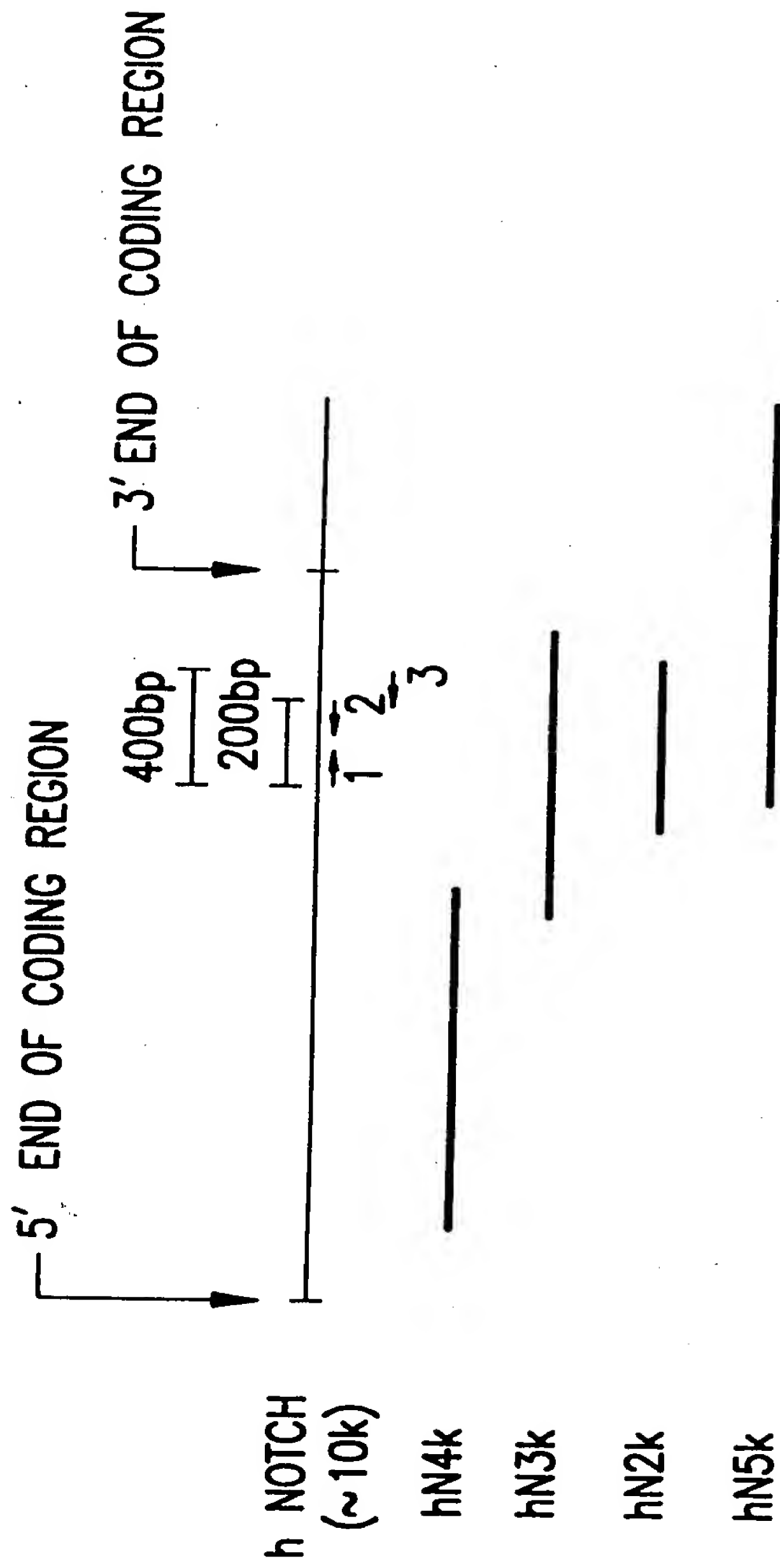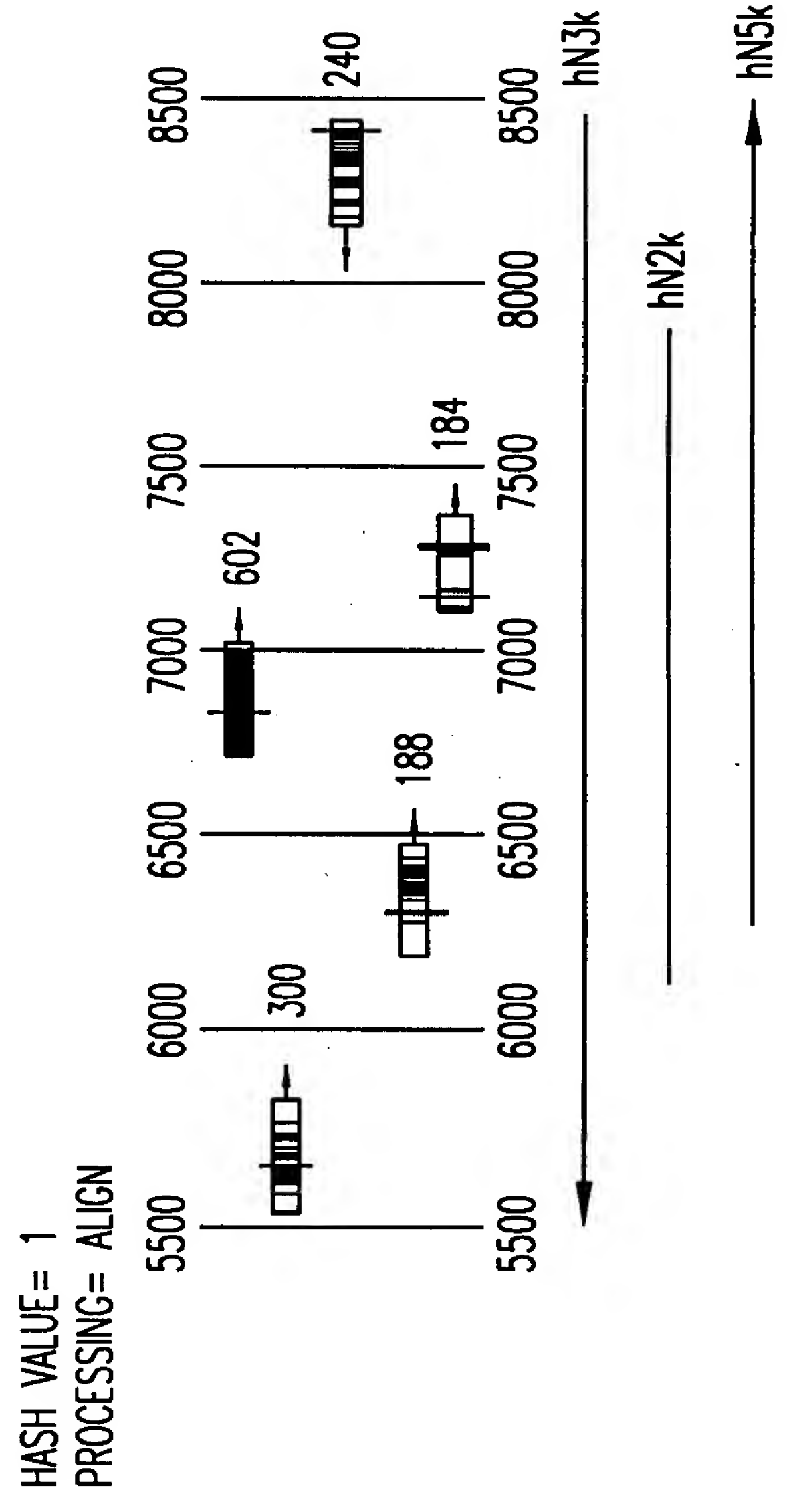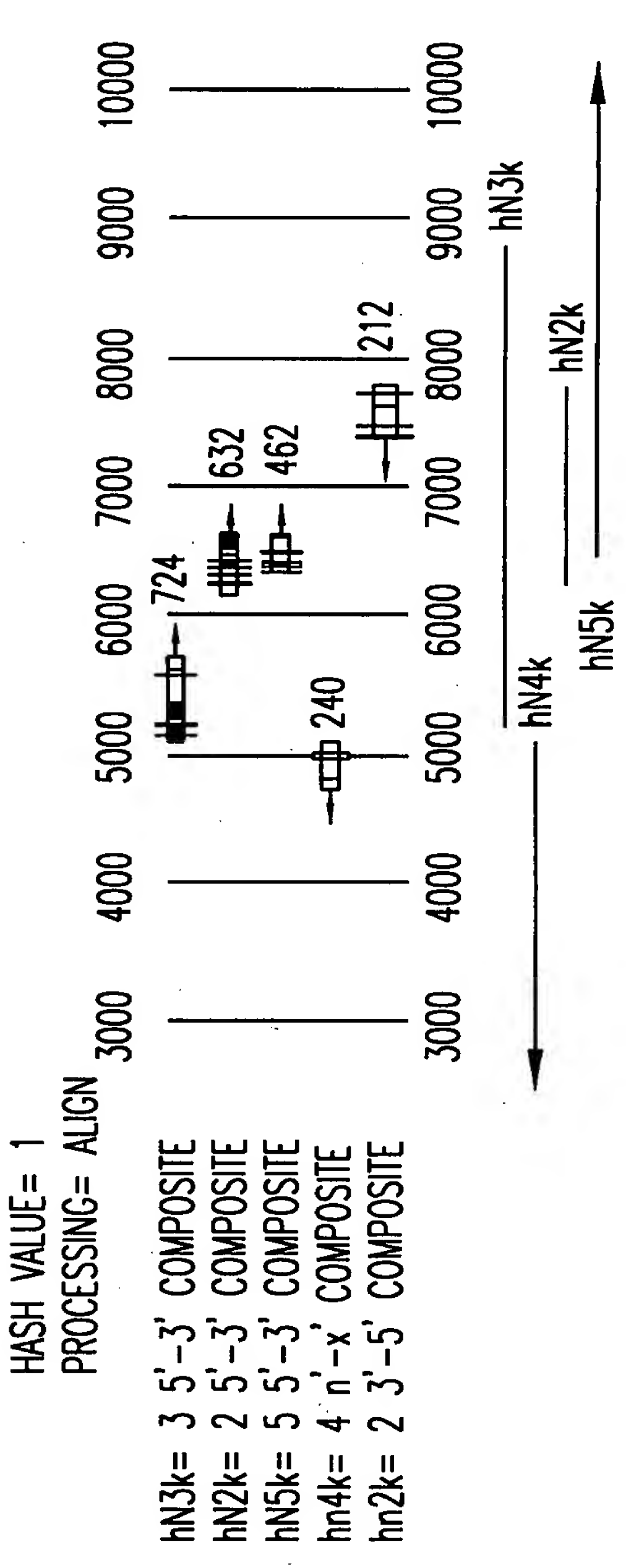hn2k= 2 3'-5' COMPOSITE

HASH VALUE= 1
PROCESSING= ALIGN

```
  1  GAATTCCGCT GGGAGAATGG TCTGAGCTAC CTGCCCGTCC TGCTGGGGCA TCAATGGCAA
 61  GTGGGGAAAG CCACACTGGG CAAACGGGCC AGGCCATTTC TGGAATGTGG TACATGGTGG
121  GCAGGGGGCC CGCAACAGCT GGAGGGCAGG TGGACTGAGG CTGGGGATCC CCCGCTGGTT
181  GGGCAATACT GCCTTTACCC ATGAGCTGGA AAGTCACAAT GGGGGGCAAG GGCTCCCGAG
241  GGTGGTTATG TGCTTCCTTC AGGTGGC
```

FIG.8A

```
  1  GAATTCCTTC CATTATACGT GACTTTCTG AAACTGTAGC CACCCTAGTG TCTCTAACTC
 61  CCTCTGGAGT TTGTCAGCTT TGGTCTTTTC AAAGAGCAGG CTCTCTTCAA GCTCCTTAAT
121  GCGGGCATGC TCCAGTTTGG TCTGCGTCTC AAGATCACCT TTGGTAATTG ATTCTTCTTC
181  AACCCGGAAC TGAAGGCTGG CTCTCACCCT CTAGGCAGAG CAGGAATTCC GAGGTGGATG
241  TGTTAGATGT GAATGTCCGT GGCCCAGATG CTGCACCCC ATTGATGTTG GCTTCTCTCC
301  GAGGAGGCAG CTCAGATTTG AGTGATGAAG ATGAAGATGC AGAGGACTGT TCTGCTAACA
361  TCATCACAGA CTTGGTCTAC CAGGGTGCCA GCCTCCAGAC CAGACAGACC GGACTGGTGA
421  GATGGCCCTG CACCTTGCAG CCCGCTACTC ACGGGCTGAT GCTGCCAAGC GTCTCCTGGA
481  TGCAGGTGCA GATGCCAATG CCCAGGACAA CATGGGCCGC TGTCCACTCC ATGCTGCAGT
541  GGCACGTGAT GCCAAGGTGT ATTCAGATCT GTTA
```

FIG.8B

```
  1  TCCAGATTCT GATTCGCAAC CGAGTAACTG ATCTAGATGC CAGGATGAAT GATGGTACTA
 61  CACCCCTGAT CCTGGCTGCC CGCCTGGCTG TGGAGGGAAT GGTGGCAGAA CTGATCAACT
121  GCCAAGCGGA TGTGAATGCA GTGGATGACC ATGGAAAATC TGCTCTTCAC TGGGCAGCTG
181  CTGTCAATAA TGTGGAGGCA ACTCTTTTGT TGTTGAAAAA TGGGGCCAAC CGAGACATGC
241  AGGACAACAA GGAAGAGACA CCTCTGTTTC TTGCTGCCCG GGAGGAGCTA TAAGC
```

FIG.8C

```
  1  GAATTCCATT CAGGAGGAAA GGGTGGGGAG AGAAGCAGGC ACCCACTTTC CCGTGGCTGG
 61  ACTCGTTCCC AGGTGGCTCC ACCGGCAGCT GTGACCGCCG CAGGTGGGGG CGGAGTGCCA
121  TTCAGAAAAT TCCAGAAAAG CCCTACCCCA ACTCGGACGG CAACGTCACA CCCGTGGGTA
181  GCAACTGGCA CACAAACAGC CAGCGTGTCT GGGGCACGGG GGGATGGCAC CCCCTGCAGG
241  CAGAGCTG
```

FIG.9A


```
  1  CTAAAGGGAA CAAAAGCNGG AGCTCCACCG CGGGCGGCNC NGCTCTAGAA CTAGTGGANN
 61  NCCCGGGCTG CAGGAATTCC GGCGGACTGG GCTCGGGCTC AGAGCGGCGC TGTGGAAGAG
121  ATTCTAGACC GGGAGAACAA GCGAATGGCT GACAGCTGGC CTCCAAAGTC ACCAGGCTCA
181  AATCGCTCGC CCTGGACATC GAGGGATGCA GAGGATCAGA ACCGGTACCT GGATGGCATG
241  ACTCGGATTT ACAAGCATGA CCAGCCTGCT TACAGGGAGC GTGANNTTTT CACATGCAGT
301  CGACAGACAC GAGCTCTATG CAT
```

FIG.9B

TGC CAG GAG GAC GCG GGC AAC AAG GTC TGC AGC CTG CAG TGC AAC AAC
 C   Q   E   D   A   G   N   K   V   C   S   L   Q   C   N   N>
        10   20   30   40

CAC GCG TGC TGC GGC TGG GAC GGT GAC TGC TCC CTC AAC TTC AAT GAC
 H   A   C   C   G   W   D   G   D   C   S   L   N   F   N   D>
    50   60   70   80   90

CCC TGG AAG AAC TGC ACG CAG TCT CTG CAG TGC TGG AAG TAC TTC AGT
 P   W   K   N   C   T   Q   S   L   Q   C   W   K   Y   F   S>
   100   110   120   130   140

GAC GGC CAC TGT GAC AGC CAG TGC AAC TCA GCC GGC TGC CTC TTC GAC
 D   G   H   C   D   S   Q   C   N   S   A   G   C   L   F   D>
   150   160   170   180   190

FIG. 10A

```
         200         210         220         230         240
 *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
GGC  TTT  GAC  TGC  CAG  CGT  GCG  GAA  GGC  CAG  TGC  AAC  CCC  CTG  TAC  GAC
 G    F    D    C    Q    R    A    E    G    Q    C    N    P    L    Y    D>

         250         260         270         280
 *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
CAG  TAC  TGC  AAG  GAC  CAC  TTC  AGC  GAC  GGG  CAC  TGC  GAC  CAG  GGC  TGC
 Q    Y    C    K    D    H    F    S    D    G    H    C    D    Q    G    C>

         290         300         310         320         330
 *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
AAC  AGC  GCG  GAG  GAG  TGC  GAG  TGG  GAC  GGG  CTG  GAC  TGT  GCG  GAG  CAT  GTA
 N    S    A    E    E    C    E    W    D    G    L    D    C    A    E    H    V>

         340         350         360         370         380
 *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
CCC  GAG  AGG  CTG  GCG  GCC  GGC  ACG  CTG  GTG  GTG  CTG  ATG  CCG
 P    E    R    L    A    A    G    T    L    V    V    L    M    P>
```

FIG.10B

18

CCG GAG CAG CTG CGC AAC AGC TCC TTC CAC TTC CTG CGG GAG CTC AGC
 P   E   Q   L   R   N   S   S   F   H   F   L   R   E   L   S>

CGC GTG CTG CAC ACC AAC GTG GTC TTC AAG CGT GAC GCA CAC GGC CAG
 R   V   L   H   T   N   V   V   F   K   R   D   A   H   G   Q>

CAG ATG ATC TTC CCC TAC TAC GGC CGC GAG GAG CTG CGC AAG CAC
 Q   M   I   F   P   Y   Y   G   R   E   E   L   R   K   H>

CCC ATC AAG CGT GCC GCC GAG GGC TGG GCC GCA CCT GAC GCC CTG CTG
 P   I   K   R   A   A   E   G   W   A   A   P   D   A   L   L>

FIG. 10C

19

580   590   600   610   620
GGC CAG GTG AAG GCC TCG CTG CCT GGT GGC AGC GAG GGT GGG CGG
 G   Q   V   K   A   S   L   P   G   G   S   E   G   G   R>

   630   640   650   660   670
CGG CGG AGG GAG CTG GAC CCC ATG GAC GTC CGC GGC TCC ATC GTC TAC
 R   R   R   E   L   D   P   M   D   V   R   G   S   I   V   Y>

   680   690   700   710   720
CTG GAG ATT GAC AAC CGG CAG TGT GTG CAG GCC TCC TCG CAG TGC TTC
 L   E   I   D   N   R   Q   C   V   Q   A   S   S   Q   C   E>

   730   740   750   760
CAG AGT GCC ACC GAC GTG GCC GCA TTC CTG GGA GCG CTC GCC TCG CTG
 Q   S   A   T   D   V   A   A   F   L   G   A   L   A   S   L>

FIG. 10D

```
                    780                 790                 800                 810
 *          *          *          *          *          *          *          *          *
GGC AGC CTC AAC ATC CCC TAC AAG ATC GAG GCC GTG CAG AGT GAG ACC
 G   S   L   N   I   P   Y   K   I   E   A   V   Q   S   E   T >-

        820                 830                 840                 850                 860
   *          *          *          *          *          *          *          *
GTG GAG CCG CCC CCG CCG CAG CTG CAC TTC ATG TAC GTG GCG GCG
 V   E   P   P   P   P   Q   L   H   F   M   Y   V   A   A >

        870                 880                 890                 900                 910
   *          *          *          *          *          *          *          *
GCC GCC TTT GTG CTT CTG TTC TTC GTG GGC TGC GGG GTG CTG CTG TCC
 A   A   F   V   L   L   F   F   V   G   C   G   V   L   L   S >

        920                 930                 940                 950                 960
   *          *          *          *          *          *          *          *
CGC AAG CGC CGG CGG CAG CAT GGC CAG CTC TGG TTC CCT GAG GGC TTC
 R   K   R   R   R   Q   H   G   Q   L   W   F   P   E   G   F >
```

FIG.10E

21

```
                    970          980          990         1000
          *     *       *     *       *     *       *     *       *
AAA GTG TCT GAG GCC AGC AAG AAG CGG CGG GAG CCC CTC GGC GAG
 K   V   S   E   A   S   K   K   R   R   E   P   L   G   E>

    1010         1020         1030         1040         1050
  *     *       *     *       *     *       *     *       *     *
GAC TCC GTG GGC CTC AAG CCC CTG AAG AAC GCT TCA GAC GGT GCC CTC
 D   S   V   G   L   K   P   L   K   N   A   S   D   G   A   L>

    1060         1070         1080         1090         1100
  *     *       *     *       *     *       *     *       *     *
ATG GAC GAC AAC CAG AAT GAG TGG GGG GAC GAG GAC CTG GAG ACC AAG
 M   D   D   N   Q   N   E   W   G   D   E   D   L   E   T   K>

    1110         1120         1130         1140         1150
  *     *       *     *       *     *       *     *       *
AAG TTC CGG TTC GAG GAG CCC GTG CTG CCT GAC CTG GAC GAC CAG
 K   F   R   F   E   E   P   V   L   P   D   L   D   D   Q>
```

FIG.10F

22

1160   1170   1180   1190   1200

ACA GAC CAC CGG CAG TGG ACT CAG CAC CTG GAT GCC GCT GAC CTG
 T   D   H   R   Q   W   T   Q   H   L   D   A   A   D   L>

      1210   1220   1230   1240

CGC ATG TCT GCC ATG GCC CCC ACA CCG CCC CAG GGT GAG GTT GAC GCC
 R   M   S   A   M   A   P   T   P   P   Q   G   E   V   D   A>

1250   1260   1270   1280   1290

GAC TGC ATG GAC GTC AAT GTC CGC GGG CCT GAT GGC TTC ACC CCG CTC
 D   C   M   D   V   N   V   R   G   P   D   G   F   T   P   L>

1300   1310   1320   1330   1340

ATG ATC GCC TCC TGC AGC GGG GGC GGC CTG GAG ACG GGC AAC AGC GAG
 M   I   A   S   C   S   G   G   G   L   E   T   G   N   S   E>

FIG.10G

1350 1360 1370 1380 1390
GAA GAG GAC GCG CCG GCC GTC ATC TCC GAC TTC ATC TAC CAG GGC
 E   E   D   A   P   A   V   I   S   D   F   I   Y   Q   G>

1400 1410 1420 1430 1440
GCC AGC CTG CAC AAC CAG ACA GAC CGC ACG GGC GAG ACC GCC TTG CAC
 A   S   L   H   N   Q   T   D   R   T   G   E   T   A   L   H>

1450 1460 1470 1480
CTG GCC CGC TAC TCA CGC TCT GAT GCC GCC AAG CGC CTG CTG GAG
 L   A   R   Y   S   R   S   D   A   A   K   R   L   L   E>

1490 1500 1510 1520 1530
GCC AGC GCA GAT GCC AAC ATC CAG GAC AAC ATG GGC CGC ACC CCG CTG
 A   S   A   D   A   N   I   Q   D   N   M   G   R   T   P   L>

FIG. 10H

1540   1550   1560   1570   1580

```
CAT GCG GCT GTG TCT GCC GAC GCA CAA GGT GTC TTC CAG ATC CTG ATC
 H   A   A   V   S   A   D   A   Q   G   V   F   Q   I   L   I>
```

1590   1600   1610   1620   1630

```
CGG AAC CGA GCC ACA GAC CTG GAT GCC CGC ATG CAT GAT GGC ACG ACG
 R   N   R   A   T   D   L   D   A   R   M   H   D   G   T   T>
```

1640   1650   1660   1670   1680

```
CCA CTG ATC CTG GCT GCC CGC CTG GCC GTG GAG GGC ATG CTG GAG GAC
 P   L   I   L   A   A   R   L   A   V   E   G   M   L   E   D>
```

1690   1700   1710   1720

```
CTC ATC AAC TCA CAC GCC GAC GTC AAC GCC GTA GAT GAC CTG GGC AAG
 L   I   N   S   H   A   D   V   N   A   V   D   D   L   G   K>
```

1730   1740   1750   1760   1770

```
TCC GCC CTG CAC TGG GCC GCC GCC GTG AAC AAT GTG GAT GCC GCA GTT
 S   A   L   H   W   A   A   A   V   N   N   V   D   A   A   V>
```

FIG.10I

1780   1790   1800   1810   1820
*   *   *   *   *   *   *
GTG CTC CTG AAG AAC GGG GCT AAC AAA GAT ATG CAG AAC AAC AGG GAG
 V   L   L   K   N   G   A   N   K   D   M   Q   N   N   R   E >

1830   1840   1850   1860   1870
*   *   *   *   *   *   *
GAG ACA CCC CTG TTT CTG GCC GCC CGG GAG GGC AGC TAC GAG ACC GCC
 E   T   P   L   F   L   A   A   R   E   G   S   Y   E   T   A >

1880   1890   1900   1910   1920
*   *   *   *   *   *   *
AAG GTG CTG CTG GAC CAC TTT GCC AAC CGG GAC ATC ACG GAT CAT ATG
 K   V   L   L   D   H   F   A   N   R   D   I   T   D   H   M >

1930   1940   1950   1960
*   *   *   *   *   *
GAC CGC CTG CCG CGC GAC ATC GCA CAG GAG CGC ATG CAT CAC GAC ATC
 D   R   L   P   R   D   I   A   Q   E   R   M   H   H   D   I >

FIG.10J

26

```
1970        1980        1990        2000        2010
  *    *      *      *      *      *      *      *    *
GTG AGG CTG CTG GAC GAG TAC AAC CTG GTG CGC AGC CCG CAG CTG CAC
 V   R   L   L   D   E   Y   N   L   V   R   S   P   Q   L   H>

2020        2030        2040        2050        2060
  *    *      *      *      *      *      *      *    *
GGA GCC CCG CTG GGG GGC ACG CCC ACC CTG TCG CCC CCG CTC TGC TCG
 G   A   P   L   G   G   T   P   T   L   S   P   P   L   C   S>

2070        2080        2090        2100        2110
  *    *      *      *      *      *      *      *    *
CCC AAC GGC TAC CTG GGC AGC CTC AAG CCC GGC GTG CAG GGC AAG AAG
 P   N   G   Y   L   G   S   L   K   P   G   V   Q   G   K   K>

2120        2130        2140        2150        2160
  *    *      *      *      *      *      *      *    *
GTC CGC AAG CCC AGC AGC AAA GGC CTG GCC TGT GGA AGC AAG GAG GCC
 V   R   K   P   S   S   K   G   L   A   C   G   S   K   E   A>
```

FIG.10K

27

```
       2170      2180      2190      2200
        *         *         *         *
AAG GAC CTC AAG GCA CGG AGG AAG AAG TCC CAG GAT GGC AAG GGC TGC
 K   D   L   K   A   R   R   K   K   S   Q   D   G   K   G   C>

 2210      2220      2230      2240      2250
  *         *         *         *         *
CTG CTG GAC AGC TCC GGC ATG CTC TCG CCC GTG GAC TCC CTG GAG TCA
 L   L   D   S   S   G   M   L   S   P   V   D   S   L   E   S>

 2260      2270      2280      2290      2300
  *         *         *         *         *
CCC CAT GGC TAC CTG TCA GAC GTG GCC TCG CCG CCA CTG CCC CTC TCC
 P   H   G   Y   L   S   D   V   A   S   P   P   L   P   L   S>

 2310      2320      2330      2340      2350
  *         *         *         *         *
CCG TTC CAG CAG TCT CCG TCC GTG CCC CTC AAC CAC CTG CCT GGG ATG
 P   F   Q   Q   S   P   S   V   P   L   N   H   L   P   G   M>
```

FIG.10L

28

2360   2370   2380   2390   2400

```
CCC GAC ACC CAC CTG GGC ATC GGG CAC CTG AAC GTG GCG GCC AAG CCC
 P   D   T   H   L   G   I   G   H   L   N   V   A   A   K   P>
```

2410   2420   2430   2440

```
GAG ATG GCG GCG CTG GCG GGT GGG GGC CGG CTG GCC TTT GAG ACT GGC
 E   M   A   A   L   G   G   G   G   R   L   A   F   E   T   G>
```

2450   2460   2470   2480   2490

```
CCA CCT CGT CTC TCC CAC CTG CCT GTG GCC TCT GGC ACC AGC ACC GTC
 P   P   R   L   S   H   L   P   V   A   S   G   T   S   T   V>
```

2500   2510   2520   2530   2540

```
CTG GGC TCC AGC AGC GGA GGG GCC CTG AAT TTC ACT GTG GGC GGG TCC
 L   G   S   S   S   G   G   A   L   N   F   T   V   G   G   S>
```

FIG.10M

```
  2550         2560         2570         2580         2590
*      *            *      *            *      *            *      *            *      *
ACC AGT TTG AAT GGT CAA TGC GAG TGG CTG TCC CGG CTG CAG AGC GGC
 T   S   L   N   G   Q   C   E   W   L   S   R   L   Q   S   G>

       2600         2610         2620         2630         2640
*      *            *      *            *      *            *      *            *      *
ATG GTG CCG AAC CAA TAC AAC CCT CTG CGG GGG AGT GTG GCA CCA GGC
 M   V   P   N   Q   Y   N   P   L   R   G   S   V   A   P   G>

       2650         2660         2670         2680
*      *            *      *            *      *            *      *
CCC CTG AGC ACA CAG GCC CCC TCC CTG CAG CAT GGC ATG GTA GGC CCG
 P   L   S   T   Q   A   P   S   L   Q   H   G   M   V   G   P>

  2690         2700         2710         2720         2730
*      *            *      *            *      *            *      *
CTG CAC AGT AGC CTT GCT GCC AGC GCC CTG TCC CAG ATG ATG AGC TAC
 L   H   S   S   L   A   A   S   A   L   S   Q   M   M   S   Y>
```

FIG.10N

```
      2740            2750            2760            2770            2780
CAG GGC CTG CCC AGC ACC CGG CTG GCC ACC CAG CCT CAC CTG GTG CAG
 Q   G   L   P   S   T   R   L   A   T   Q   P   H   L   V   Q>

      2790            2800            2810            2820            2830
ACC CAG CAG GTG CAG CCA CAA AAC TTA CAG ATG CAG CAG AAC CTG
 T   Q   Q   V   Q   P   Q   N   L   Q   M   Q   Q   N   L>

      2840            2850            2860            2870            2880
CAG CCA GCA AAC ATC CAG CAG CAG CAA AGC CTG CAG CCG CCA CCA
 Q   P   A   N   I   Q   Q   Q   Q   S   L   Q   P   P   P>

      2890            2900            2910            2920
CCA CCA CAG CCG CAC CTT GGC GTG AGC TCA GCA GCC AGC GGC CAC CTG
 P   P   Q   P   H   L   G   V   S   S   A   A   S   G   H   L>

      2930            2940            2950            2960            2970
GGC CGG AGC TTC CTG AGT GGA GAG CCG AGC CAG GCA GAC GTG CAG CCA
 G   R   S   F   L   S   G   E   P   S   Q   A   D   V   Q   P>
```

FIG.100

31

2980   2990   3000   3010   3020

CTG GGC CCC AGC CTG GCG GTG CAC ACT ATT CTG CCC CAG GAG AGC
L   G   P   S   L   A   V   H   T   I   L   P   Q   E   S>

3030   3040   3050   3060   3070

CCC GCC CTG CCC ACG TCG CTG CCA TCC TCG CTG GTC CCA CCC GTG ACC
P   A   L   P   T   S   L   P   S   S   L   V   P   P   V   T>

3080   3090   3100   3110   3120

GCA GCC CAG TTC CTG ACG CCC CCC TCG CAG CAC AGC TAC TCC TCG CCT
A   A   Q   F   L   T   P   P   S   Q   H   S   Y   S   S   P>

FIG.10P

32

```
        3130            3140            3150            3160
GTG GAC AAC ACC CCC AGC CAC CAG CTA CAG GTG CCT GTT CCT GTA ATG
 V   D   N   T   P   S   H   Q   L   Q   V   P   V   P   V   M>

        3170            3180            3190            3200            3210
GTA ATG ATC CGA TCT TCG GAT CCT TCT AAA GGC TCA TCA ATT TTG ATC
 V   M   I   R   S   S   D   P   S   K   G   S   S   I   L   I>

        3220            3230
GAA GCT CCC GAC TCA TGG
 E   A   P   D   S   W>
```

FIG.10Q

```
G GAG GTG GAT GTG TTA GAT GTG AAT GTC CGT GGC CCA GAT GGC TGC        46
  Glu Val Asp Val Leu Asp Val Asn Val Arg Gly Pro Asp Gly Cys
  1               5                   10                  15


ACC CCA TTG ATG TTG GCT TCT CTC CGA GGA GGC AGC TCA GAT TTG AGT      94
Thr Pro Leu Met Leu Ala Ser Leu Arg Gly Gly Ser Ser Asp Leu Ser
            20                  25                  50


GAT GAA GAT GAA GAT GCA GAG GAC TCT TCT GCT AAC ATC ATC ACA GAC      142
Asp Glu Asp Glu Asp Ala Glu Asp Ser Ser Ala Asn Ile Ile Thr Asp
            35                  40                  45


TTG GTC TAC CAG GGT GCC AGC CTC CAG GCC CAG ACA GAC CGG ACT GGT      190
Leu Val Tyr Gln Gly Ala Ser Leu Gln Ala Gln Thr Asp Arg Thr Gly
        50                  55                  60


GAG ATG GCC CTG CAC CTT GCA GCC CGC TAC TCA CGG GCT GAT GCT GCC      238
Glu Met Ala Leu His Leu Ala Ala Arg Tyr Ser Arg Ala Asp Ala Ala
    65                  70                  75


AAG CGT CTC CTG GAT GCA GGT GCA GAT GCC AAT GCC CAG GAC AAC ATG      286
Lys Arg Leu Leu Asp Ala Gly Ala Asp Ala Asn Ala Gln Asp Asn Met
80                  85                  90                  95


GGC CGC TGT CCA CTC CAT GCT GCA GTG GCA GCT GAT GCC CAA GGT GTC      334
Gly Arg Cys Pro Leu His Ala Ala Val Ala Ala Asp Ala Gln Gly Val
            100                 105                 110


TTC CAG ATT CTG ATT CGC AAC CGA GTA ACT GAT CTA GAT GCC AGG ATG      382
Phe Gln Ile Leu Ile Arg Asn Arg Val Thr Asp Leu Asp Ala Arg Met
            115                 120                 125


AAT GAT GGT ACT ACA CCC CTG ATC CTG GCT GCC CGC CTG GCT GTG GAG      430
Asn Asp Gly Thr Thr Pro Leu Ile Leu Ala Ala Arg Leu Ala Val Glu
            130                 135                 140
```

FIG.11A

GGA ATG GTG GCA GAA CTG ATC AAC TGC CAA GCG GAT GTG AAT GCA GTG   478
Gly Met Val Ala Glu Leu Ile Asn Cys Gln Ala Asp Val Asn Ala Val
     145   150   155

GAT GAC CAT GGA AAA TCT GCT CTT CAC TGG GCA GCT GCT GTC AAT AAT   526
Asp Asp His Gly Lys Ser Ala Leu His Trp Ala Ala Ala Val Asn Asn
160   165   170   175

GTG GAG GCA ACT CTT TTG TTG TTG AAA AAT GGG GCC AAC CGA GAC ATG   574
Val Glu Ala Thr Leu Leu Leu Leu Lys Asn Gly Ala Asn Arg Asp Met
             180   185   190

CAG GAC AAC AAG GAA GAG ACA CCT CTG TTT CTT GCT GCC CGG GAG GGG   622
Gln Asp Asn Lys Glu Glu Thr Pro Leu Phe Leu Ala Ala Arg Glu Gly
          195   200   205

AGC TAT GAA GCA GCC AAG ATC CTG TTA GAC CAT TTT GCC AAT CGA GAC   670
Ser Tyr Glu Ala Ala Lys Ile Leu Leu Asp His Phe Ala Asn Arg Asp
         210   215   220

ATC ACA GAC CAT ATG GAT CGT CTT CCC CGG GAT GTG GCT CGG GAT CGC   718
Ile Thr Asp His Met Asp Arg Leu Pro Arg Asp Val Ala Arg Asp Arg
    225   230   235

ATG CAC CAT GAC ATT GTG CGC CTT CTG GAT GAA TAC AAT GTG ACC CCA   766
Met His His Asp Ile Val Arg Leu Leu Asp Glu Tyr Asn Val Thr Pro
240   245   250   255

AGC CCT CCA GGC ACC GTG TTG ACT TCT GCT CTC TCA CCT GTC ATC TGT   814
Ser Pro Pro Gly Thr Val Leu Thr Ser Ala Leu Ser Pro Val Ile Cys
             260   265   270

GGG CCC AAC AGA TCT TTC CTC AGC CTG AAG CAC ACC CCA ATG GGC AAG   862
Gly Pro Asn Arg Ser Phe Leu Ser Leu Lyn His Thr Pro Met Gly Lys
         275   280   285

FIG.11B

35

```
AAG TCT AGA CGG CCC AGT GCC AAG AGT ACC ATG CCT ACT AGC CTC CCT    910
Lys Ser Arg Arg Pro Ser Ala Lys Ser Thr Met Pro Thr Ser Leu Pro
        290             295                 300

AAC CTT GCC AAG GAG GCA AAG GAT GCC AAG GGT AGT AGG AGG AAG AAG    958
Asn Leu Ala Lys Glu Ala Lys Asp Ala Lys Gly Ser Arg Arg Lys Lys
        305             310                 315

TCT CTG AGT GAG AAG GTC CAA CTG TCT GAG AGT TCA GTA ACT TTA TCC   1006
Ser Leu Ser Glu Lys Val Gln Leu Ser Glu Ser Ser Val Thr Leu Ser
320             325                 330                 335

CCT GTT GAT TCC CTA GAA TCT CCT CAC ACG TAT GTT TCC GAC ACC ACA   1054
Pro Val Asp Ser Leu Glu Ser Pro His Thr Tyr Val Ser Asp Thr Thr
                340                 345                 350

TCC TCT CCA ATG ATT ACA TCC CCT GGG ATC TTA CAG GCC TCA CCC AAC   1102
Ser Ser Pro Met Ile Thr Ser Pro Gly Ile Leu Gln Ala Ser Pro Asn
                355                 360                 365

CCT ATG TTG GCC ACT GCC GCC CCT CCT GCC CCA GTC CAT GCC CAG CAT   1150
Pro Met Leu Ala Thr Ala Ala Pro Pro Ala Pro Val His Ala Gln His
        370                 375                 380

GCA CTA TCT TTT TCT AAC CTT CAT GAA ATG CAG CCT TTG GCA CAT GGG   1198
Ala Leu Ser Phe Ser Asn Leu His Glu Met Gln Pro Leu Ala His Gly
        385                 390                 395

GCC AGC ACT GTG CTT CCC TCA GTG AGC CAG TTG CTA TCC CAC CAC CAC   1246
Ala Ser Thr Val Leu Pro Ser Val Ser Gln Leu Leu Ser His His His
400             405                 410                 415

ATT GTG TCT CCA GGC AGT GGC AGT GCT GGA AGC TTG AGT AGG CTC CAT   1294
Ile Val Ser Pro Gly Ser Gly Ser Ala Gly Ser Leu Ser Arg Leu His
                420                 425                 430

CCA GTC CCA GTC CCA GCA GAT TGG ATG AAC CGC ATG GAG GTG AAT GAG   1342
Pro Val Pro Val Pro Ala Asp Trp Met Asn Arg Met Glu Val Asn Glu
                435                 440                 445
```

FIG.11C

36

ACC CAG TAC AAT GAG ATG TTT GGT ATG GTC CTG GCT CCA GCT GAG GGC 1390
Thr Gln Tyr Asn Glu Met Phe Gly Met Val Leu Ala Pro Ala Glu Gly
    450   455   460

ACC CAT CCT GGC ATA GCT CCC CAG AGC AGG CCA CCT GAA GGG AAG CAC 1438
Thr His Pro Gly Ile Ala Pro Gln Ser Arg Pro Pro Glu Gly Lys His
    465   470   475

ATA ACC ACC CCT CGG GAG CCC TTG CCC CCC ATT GTG ACT TTC CAG CTC 1486
Ile Thr Thr Pro Arg Glu Pro Leu Pro Pro Ile Val Thr Phe Gln Leu
480   485   490   495

ATC CCT AAA GGC AGT ATT GCC CAA CCA GCG GGG GCT CCC CAG CCT CAG 1534
Ile Pro Lys Gly Ser Ile Ala Gln Pro Ala Gly Ala Pro Gln Pro Gln
                500   505   510

TCC ACC TGC CCT CCA GCT GTT GCG GGC CCC CTG CCC ACC ATG TAC CAG 1582
Ser Thr Cys Pro Pro Ala Val Ala Gly Pro Leu Pro Thr Met Tyr Gln
                515   520   525

ATT CCA GAA ATG GCC CGT TTG CCC AGT GTG GCT TTC CCC ACT GCC ATG 1630
Ile Pro Glu Met Ala Arg Leu Pro Ser Val Ala Phe Pro Thr Ala Met
        530   535   540

ATG CCC CAG CAG GAC GGG CAG GTA GCT CAG ACC ATT CTC CCA GCC TAT 1678
Met Pro Gln Gln Asp Gly Gln Val Ala Gln Thr Ile Leu Pro Ala Tyr
    545   550   555

CAT CCT TTC CCA GCC TCT GTG GGC AAG TAC CCC ACA CCC CCT TCA CAG 1726
His Pro Phe Pro Ala Ser Val Gly Lys Tyr Pro Thr Pro Pro Ser Gln
560   565   570   575

CAC AGT TAT GCT TCC TCA AAT GCT GCT GAG CGA ACA CCC AGT CAC AGT 1774
His Ser Tyr Ala Ser Ser Asn Ala Ala Glu Arg Thr Pro Ser His Ser
                580   585   590

GGT CAC CTC CAG GGT GAG CAT CCC TAC CTG ACA CCA TCC CCA GAG TCT 1822
Gly His Leu Gln Gly Glu His Pro Tyr Leu Thr Pro Ser Pro Glu Ser
            595   600   605

FIG.11D

37

```
CCT GAC CAG TGG TCA AGT TCA TCA CCC CAC TCT GCT TCT GAC TGG TCA        1870
Pro Asp Gln Trp Ser Ser Ser Ser Pro His Ser Ala Ser Asp Trp Ser
        610                 615                 620

GAT GTG ACC ACC AGC CCT ACC CCT GGG GGT GCT GGA GGA GGT CAG CGG        1918
Asp Val Thr Thr Ser Pro Thr Pro Gly Gly Ala Gly Gly Gly Gln Arg
        625                 630                 635

GGA CCT GGG ACA CAC ATG TCT GAG CCA CCA CAC AAC AAC ATG CAG GTT        1966
Gly Pro Gly Thr His Met Ser Glu Pro Pro His Asn Asn Met Gln Val
640                 645                 650                 655

TAT GCG TGAGAGAGTC CACCTCCAGT GTAGAGACAT AACTGACTTT TGTAAATGCT         2022
Tyr Ala


GCTGAGGAAC AAATGAAGGT CATCCGGGAG AGAAATGAAG AAATCTCTGG AGCCAGCTTC      2082

TAGAGGTAGG AAAGAGAAGA TGTTCTTATT CAGATAATGC AAGAGAAGCA ATTCGTCAGT      2142

TTCACTGGGT ATCTGCAAGG CTTATTGATT ATTCTAATCT AATAAGACAA GTTTGTGGAA      2202

ATGCAAGATG AATACAAGCC TTGGGTCCAT GTTACTCTC TTCTATTTGG AGAATAAGAT       2262

GGATGCTTAT TGAAGCCCAG ACATTCTTGC AGCTTGGACT GCATTTTAAG CCCTGCAGGC      2322

TTCTGCCATA TCCATGAGAA GATTCTACAC TAGCGTCCTG TTGGGAATTA TGCCCTGGAA      2382

TTCTGCCTGA ATTGACCTAC GCATCTCCTC CTCCTTGGAC ATTCTTTTGT CTTCATTTGG      2442

TGCTTTTGGT TTTGCACCTC TCCGTGATTG TAGCCCTACC AGCATGTTAT AGGGCAAGAC      2502

CTTTGTGCTT TTGATCATTC TGGCCCATGA AAGCAACTTT GGTCTCCTTT CCCCTCCTGT      2562

CTTCCCGGTA TCCCTTGGAG TCTCACAAGG TTTACTTTGG TATGGTTCTC AGCACAAACC      2622

TTTCAAGTAT GTTGTTTCTT TGGAAAATGG ACATACTGTA TTGTGTTCTC CTGCATATAT      2682

CATTCCTGGA GAGAGAAGGG GAGAAGAATA CTTTTCTTCA ACAAATTTTG GGGGCAGGAG      2742

ATCCCTTCAA GAGGCTGCAC CTTAATTTTT CTTGTCTGTG TGCAGGTCTT CATATAAACT      2802
```

FIG.11E

38

```
TTACCAGGAA GAAGGGTGTG AGTTTGTTGT TTTTCTGTGT ATGGGCCTGG TCAGTGTAAA    2862

GTTTTATCCT TGATAGTCTA GTTACTATGA CCCTCCCCAC TTTTTTAAAA CCAGAAAAAG    2922

GTTTGGAATG TTGGAATGAC CAAGAGACAA GTTAACTCGT GCAAGAGCCA GTTACCCACC    2982

CACAGGTCCC CCTACTTCCT GCCAAGCATT CCATTGACTG CCTGTATGGA ACACATTTGT    3042

CCCAGATCTG AGCATTCTAG GCCTGTTTCA CTCACTCACC CAGCATATGA AACTAGTCTT    3102

AACTGTTGAG CCTTTCCTTT CATATCCACA GAAGACACTG TCTCAAATGT TGTACCCTTG    3162

CCATTTAGGA CTGAACTTTC CTTAGCCCAA GGGACCCAGT GACAGTTGTC TTCCGTTTGT    3222

CAGATGATCA GTCTCTACTG ATTATCTTGC TGCTTAAAGG CCTGCTCACC AATCTTTCTT    3282

TCACACCGTG TGGTCCGTGT TACTGGTATA CCCAGTATGT TCTCACTGAA GACATGGACT    3342

TTATATGTTC AAGTGCAGGA ATTGGAAAGT TGGACTTGTT TTCTATGATC CAAAACAGCC    3402

CTATAAGAAG GTTGGAAAAG GAGGAACTAT ATAGCAGCCT TTGCTATTTT CTGCTACCAT    3462

TTCTTTTCCT CTGAAGCGGC CATGACATTC CCTTTGGCAA CTAACGTAGA AACTCAACAG    3522
```

FIG.11F

39

```
AACATTTTCC TTTCCTAGAG TCACCTTTTA GATGATAATG GACAACTATA GACTTGCTCA    3582

TTGTTCAGAC TGATTGCCCC TCACCTGAAT CCACTCTCTG TATTCATGCT CTTGGCAATT    3642

TCTTTGACTT TCTTTTAAGG GCAGAAGCAT TTTAGTTAAT TGTAGATAAA GAATAGTTTT    3702

CTTCCTCTTC TCCTTGGGCC AGTTAATAAT TGGTCCATGG CTACACTGCA ACTTCCGTCC    3762

AGTGCTGTGA TGCCCATGAC ACCTGCAAAA TAAGTTCTGC CTGGGCATTT TGTAGATATT    3822

AACAGGTGAA TTCCCGACTC TTTTGGTTTG AATGACAGTT CTCATTCCTT CTATGGCTGC    3882

AAGTATGCAT CAGTGCTTCC CACTTACCTG ATTTGTCTGT CGGTGGCCCC ATATGGAAAC    3942

CCTGCGTGTC TGTTGGCATA ATAGTTTACA AATGGTTTTT TCAGTCCTAT CCAAATTTAT    4002

TGAACCAACA AAAATAATTA CTTCTGCCCT GAGATAAGCA GATTAAGTTT GTTCATTCTC    4062

TGCTTTATTC TCTCCATGTG GCAACATTCT GTCAGCCTCT TTCATAGTGT GCAAACATTT    4122

TATCATTCTA AATGGTGACT CTCTGCCCTT GGACCCATTT ATTATTCACA GATGGGGAGA    4182

ACCTATCTGC ATGGACCCTC ACCATCCTCT GTGCAGCACA CACAGTGCAG GGAGCCAGTG    4242

GCGATGGCGA TGACTTTCTT CCCCTG                                         4268
```
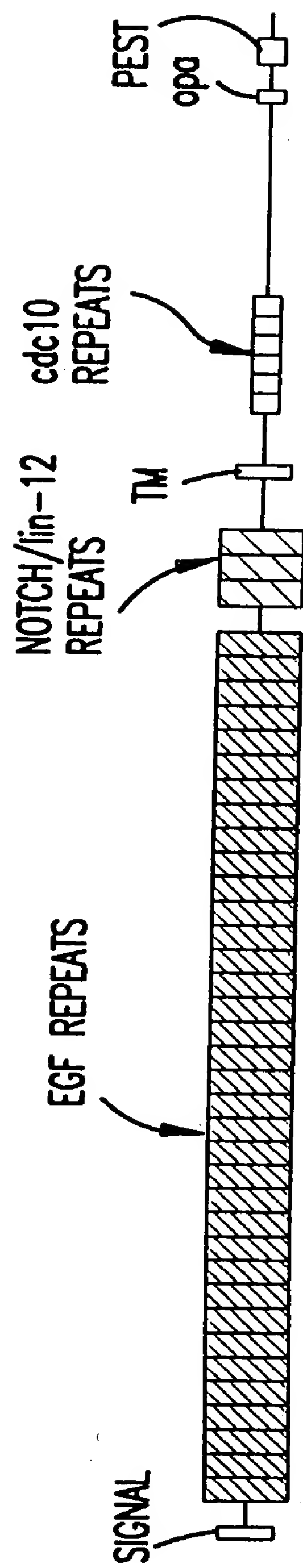
FIG. 11G

FIG.12A

CDC10 REGION

CDC10-1 →   CDC10-2 →

hN5k	TPPQAEQEVDVL--DVNVRGPDGCTPLMLASLRGGSSDL--S-DEDEADEDSSANIITDLVYQGASLQAQTDRTGEMALHLAARYSRAIDAAKRLLDAGA	1860
TAN-1	TPPQG--EVDADCMDVNVRGPDGFTPLMIASCSGGGLETGNS-EEEEDAP---A-VISDFIYQGASLHNQTDRTGETALHLAARYSRSIDAAKRLLEASA	1857
rat NOTCH	TPPQG--EVDADCMDVNVRGPDGFTPLMIASCSGGGLETGNS-EEEEDAP---A-VISDFIYQGASLHNQTDRTGETALHLAARYSRSIDAAKRLLEASA	1855
XENOPUS NOTCH	TPPQG--EIEADCMDVNVRGPDGFTPLMIASCSGGGLETGNS-EEEEDA--S-ANMISDFIGQGAQLHNQTDRTGETALHLAARYARAIDAAKRLLESSA	1883
DROSOPH NOTCH	TPPA---HQDGGKHDVDARGPCGLTPLMIAAVRGGGLDTGEDIENNED---STAQVISDLLAQGAELNATMDKTGETSLHLAARFARAIDAAKRLFHAGA	

CDC10-3 →   CDC10-4 →   CDC10-5 →   CDC10-6 →

hN5k	DANAQDNMGRCPLHAAVAADAQGVFQILIRNRVTDLDARMNDGTPLLAARLAVEGMVAELINCQADVNAVDDHGKSALHWAAAVNNVEATLLLLKNGANRDMQD	1952
TAN-1	DANIQDNMGRTPLHAAYSADAQGVFQILIRNRATDLDARMHDGTPLLAARLAVEGMLEDLINSHADVNAVDDLGKSALHWAAAVNNVDAAVVLLKNGANKDMQN	1944
rat NOTCH	DANIQDNMGRTPLHAAYSADAQGVFQILIRNRATDLDARMHDGTPLLAARLAVEGMLEDLINSHADVNAVDDLGKSALHWAAAVNNVDAAVVLLKNGANKDMQN	1950
XENOPUS NOTCH	DANVQDNMGRTPLHAAVAADAQGVFQILIRNRATDLDARMFDGTPLLAARLAVEGMVEELINAHADVNAVDEFGKSALHWAAAVNNVDAAAVLLKNSANKDMQN	1976
DROSOPH NOTCH	DANCQDNTGRTPLHAAVAADAMGVFQILLRNRATNLNARMHDGTPLLAARLAIEGMVEDLITADADRNAADNSGKTALHWAAAVNNTEAVNILLMHHANRDAQD	

PRIMER cdc-1

GAC/T GCI AAC/T GTI CAAA/G GAC/TAAC/T ATG GG

hN5k	NKEETPLFLAAREGSYEAAKILLDHFANRDITDHMDRLPRDVARDRMHHDIVRLLDEYNVTPSPP--GTVL--TSALSPV-------ICGPNRSFLSLKHTP	2058
TAN-1	NREETPLFLAAREGSYETAKVLLDHFANRDITDHMDRLPRDIAQERMHHDIVRLLDEYNLVRSPQLHGAPLGGTPTLSPP-------LCSPNGYLGSLKPGV	2050
rat NOTCH	NKEETPLFLAAREGSYETAKVLLDHFANRDITDHMDRLPRDIAQERMHHDIVRLLDEYNLVRSPQLHGTALGGTPTLSPT-------LCSPNGYLGNLKSAT	2056
XENOPUS NOTCH	NKEETSLFLAAREGSYETAKVLLDHYANRDITDHMDRLPRDIAQERMHHDIVHLLDEYNLVKSPTLHNGPLGAT-ILSPP-------ICSPNGYMGNMKPSV	2082
DROSOPH NOTCH	DKDETPLFLAAREGSYEACKALLDNFANREITDHMDRLPRDVASRLHHDIVRLLDE-HVPRSPQMLSMTPQAMIGSPPPGQQPQLITQPTVISAGNGNGNGN	

TC CAT G^ATG G^ATC ICT T^GV^AT ITC ICT^G G^ATT   PRIMER cdc-3

hN5k	MGKKSRRPSAKSTMPTSLPNLAKEAKDAKGSRRKKQSLSEKVQLSE--SSVTLSPVDSLESPHTIYSDTTSSPM------------	2154
TAN-1	QGKKVRKPSSKGLACGS----KEAKDLKA-RRKKQSDBGKGCLLD--SSGMLSPVDSLESPHGYLSDVASPPL-------------	2144
rat NOTCH	QGKKARKPSTKGLACSS----KEAKDLKA-RRKKQSDBGKGCLLD--SSSMLSPVDSLESPHGYLSDVASPPL-------------	2149
XENOPUS NOTCH	QSKKARKPSIKGNGC------KEAKELKA-RRKKQSDGKTTLLDSGSSGVLSPVDSLESTHGYLSDVSSPPL--------------	2187
DROSOPH NOTCH	ASGKQSNQTAKQKAA------KKAKLIEGS-PDNGLDATGSDRRHASSKKTSAASKKAANLNGLNPGQLTGGVSGVPGVPPTNSAQGAAAAAAVAAMSHELEGS	

NLS   CK-II   cdc2   cdc2
BNTS

FIG. 12B

42

hN5k   ------------ITSPGIDDASPNPML--ATAAPPAPVHAQHALSF
TAN-1   2218 ------------LRSPR--QQSPSVPLNHLPGMPDTHLGIGHLNVA
rat NOTCH   2209 ------------LRSPR--QQSPSMPLSHLPGMPDTHLGISHLNVA
XENOPUS NOTCH   2214 ------------MTSPR--QQSPSMPLNHLTSMPESQLGMNHINMA
DROSOPH NOTCH   2285 PVGVGMGGNLPSPYDTSSMYSNAMAAPLANGNPNTGAKQPRSYEDCIKNAQSMQSLQGNGLDMIKLDNYAYSMGSPR--QQELLNGQGLGMGNGQRNGVGPGVLP

CK-II

hN5k   SNLHEMQ------------PLAHGASTVLPSVSVQLLSHHHIVSPGS--GSAGSLSRLHPVPVPADW--MNRMEVNETQYNEMFGMVLAPAEG-THPGI
TAN-1   2250 A-KPEMAALGGGRLAFETGPPRLSHLPVASGTSTVLGSSSGGALNFTVGGSTSLNGGQCEVLSRLQSGMVPNQYNPLRGSVAPGPLSTQAPSLQHG-MVGPLHSSL
rat NOTCH   2242 A-KPEMAALAGGSRLAFEPPPRLSHLPVASSASTVLSTNGTGAMNFTVGAPASLNGQCEVLPRLQNGMVPSQYNPLRPGVTPGTLSTQAAGLQHGMM-SPIHSSL
XENOPUS NOTCH   2247 T-KQEMAA--GSNRMAFDAMVPRLTHL-NASSPNTIMS---NGSMHFTVGGAPTMNSQCDWLARLQNGMVQNQYDPIRNGIQQGN-AQQAQALQHGLMTS-LHNGL
DROSOPH NOTCH   2390 GGLCQMGGLSGAGNGNSHEQGLSPPYS-NQSPPHSVQSSLALSPHAYLGSPSPAKSRPSLPTSPTHIQAMRHATQQKQFGGSNLNSLLGGANGGGVVGGGGGGGV

CK-II

hN5k   APQSRPPEGK------------HITTPREPLPP-IV-TFQLIPKGSIAQPAG------------APQPQSTCPPAVAGPLPTMYQIP-----EMARL-P
TAN-1   2354 AASALSQMMS------------YQGLPSTRLATQPHLVQTQQVQPQNLQMQQQNLQPANIQQQQSLQPPPPPQPHLGVSSAASGHLGRSFLSGEPSQADVQPLGP
rat NOTCH   2344 SINTLSPII------------YQGLPNTRLATQPHLVQTQQVQPQNLQIQPQN------------LQPPS--QPHLSVSSAANGHLGRSFLSGEPSQADVQPLGP
XENOPUS NOTCH   2343 PATTLSQMMT------------YQAMPNTRLANQPHLMQAQQMQQQN------------LQLHQSMQQQHHN-SSTTSTHINSPFCSSDISQTDLQQM--
DROSOPH NOTCH   2495 VGQGPQNSPVSLGIISPTGSDMGIMLAPPQSSKNSAIMQTISPQQQQQQQQQQQQQQQQHQQQQQQQQQQQQLGGLEFGSAGLDLNG-FCGSPDSFHSGQMNPP

hN5k   SVAFPTAMMPQQDGQVAQTILPAYHPFPASVGKYHTPPSQHSYASSNAAERTPSHSGHLQGEHPYLTPSPESPDQWSSSSSPHSA-SDWSDVTTSPTP
TAN-1   2448 SSLAVHTILPQ-ESPALPTSLPSSLVPPVTAAQFLTPPSQHSY-SS-PVENTPSHQLQVP-EHPFLTPSPESPDQWSSSSPHSNVSDWSEGVSSPPT
rat NOTCH   2423 SSLPVHTILPQ-ESQALPTSLPSSMVPPMTTQFLTPPSQHSY-SSSPVDNTPSHQLQVP-EHPFLTPSPESPDQWSSSSSRHSNISDWSEGISSPPT
XENOPUS NOTCH   2416 SSNNIHSVMPQ-DTQIFAASLPSNLTQSMTTAQFLTPPSQHSY-SS-PMDNTPSHQLQVP-DHPFLTPSPESPDQWSSSSPHSNMSDWSEGISSPPT
DROSOPH NOTCH   2599 S---IQSSMSG-SSPSTNMLSPSSQHNQQAFYQYLTPSSQHS-----GGHTPQHLVQTL-D-SYPTPSPESPGHWSSSSPRSN-SDWSEGVQSPAA

PEST-CONTAINING REGION

FIG.12C

Potential signal cleavage site ¬

```
hum N    MP————— ————— ————ALRPAL LWALLALWLC CA———APA HA————⌐L
TAN-1    MP————— ————— ———PL LAPLLCLALL PA-———LAA RG————P
Xen N    MD————— ————— —————RIGLAVLLCS LP———VLT QG————L
Dros N   MQSQRSRRRS RAPNTWICFW INKMHAVASL PASLPLLLLT LAFANLPNIV RGTDTALVAA
```

```
hum N    MLGKATCRCA SGFTGEDCQY STSHPCFVSR PCLNGGTCHM LSRDT-YECT CQVGFTGKEC
Tan-1    GVADYACSCA LGFSGPLCLT PLDNAC-LTN PCRNGGTCDL LT-LTEYKCR CPPGWSGKSC
Xen N    NAIDFICHCP VGFTDKVCLT PVDNAC-VNN PCRNGGTCEL LNSVTEYKCR CPPGWTGDSC
Dros N   GRPGISCKCP LGFDESLCEI AVPNAC-DHV TCLNGGTCQL KT-LEEYTCA CANGYTGERC
```

```
hum N    NLPGSYQCQC PQGFTGQYCD SLYVPCAPSP CVNGGTCRQT GDFTFECNCL PGFEGSTCER
TAN-1    NEVGSYRCVC RATHTGPNCE RPYVPCSPSP CQNGGTCRPT GDVTHECACL PGFTGQNCEE
Xen N    NEFGSYRCTC QNRFTGRNCD EPYVPCNPSP CLNGGTCRQT DDTSYDCTCL PGFSGQNCEE
Dros N   NTHGSYQCMC PTGYTGKDCD TKYNPCSPSP CQNAGICRSN G-LSYECKCP KGFEGKNCEQ
```

EGF-like Repeats

```
QCRDGYEPCV NEGMCVTYHN GTGYCKCPEG FLGEYCQHRD PCE-KNRCQN GGTC—VAQA   83
RCSQPGETCL NGGKCEA-AN GTEACVCGGA FVGPRCQDPN PCL-STPCKN AGTCHVVDRR   80
RCTQTAEMCL NGGRCEMTPG GTGVCLCGNL YFGERCQFPN PCTIKNQCMN FGTCEPVLQG   90
SCTSVG—CQ NGGTCVTQLN GKTYCACDSH YVGDYCEHRN PCN-SMRCQN GGTCQVTFRN   117
```

```
QWTDACLSHP CANGSTCTTV —ANQFSCKC LTGFTGQKCE TDVNEC-DIP GHCQHGGTCL   199
QQADPCASNP CANGGQCLPF —EASYICHC PPSFHGPTCR QDVNECGQKP RLCRHGGTCH   196
QQADPCASNP CANGGKCLPF —EIQYICKC PPGFHGATCK QDINEC-S-Q NPCKNGGQCI   195
ETKNLCASSP CRNGATCTAL AGSSSFTCSC PPGFTGDTCS YDIEEC-Q-S NPCKYGGICV   233
```

```
NIDDCPNHRC QNGGVCVDGV NTYNCRCPPQ WTGQFCTEDV DECLLQPNA- CQNGGTCANR   318
NIDDCPGNNC KNGGACVDGV NTYNCPCPPE WTGQYCTEDV DECQLMPNA- CQNGGTCHNT   315
NIDDCPSNNC RNGGTCVDGV NTYNCQCPPD WTGQYCTEDV DECQLMPNA- CQNGGTCHNT   314
NYDDCLGHLC QNGGTCIDGI SDYTCRCPPN FTGRFCQDDV DECAQRDHPV CQNGATCTNT   352
```

FIG.13A

```
hum N    NGGYGCVCVN GWSGDDCSEN IDDCAFASCT PGSTCIDRVA SFSCMCPEGK AGLLCHLDDA
TAN-1    HGGYNCVCVN GWTGEDCSEN IDDCASAACF HGATCHDRVA SFYCECPHGR TGLLCHLNDA
Xen N    YGGYNCVCVN GWTGEDCSEN IDDCANAACH SGATCHDRVA SFYCECPHGR TGLLCHLDNA
Dros N   HGSYSCICVN GWAGLDCSNN TDDCKQAACF YGATCIDGVG SFYCQCTKGK TGLLCHLDDA

hum N    AFHCECLKGY AGPRCEMDIN ECHSDPCQND ATCLDKIGGF TCLCMPGFKG VHCELEINEC
TAN-1    SFECQCLQGY TGPRCEIDVN ECVSNPCQND ATCLDQIGEF QCMCMPGYEG VHCEVNTDEC
Xen N    SFQCNCPQGY AGPRCEIDVN ECLSNPCQND STCLDQIGEF QCICMPGYEG LYCETNIDEC
Dros N   SYRCNCSQGF TGPRCETNIN ECESHPCQNE GSCLDDPGTF RCVCMPGFTG TQCEIDIDEC

hum N    ATGFTGVLCE ENIDNCDPDP CHHGQCQDGI DSYTCICNPG YMGAICSDQI DECYSSPCLN
TAN-1    TEGYTGTHCE VDIDECDPDP CHYGSCKDGV ATFTCLCRPG YTGHHCETNI NECSSQPCRL
Xen N    TEGFTGRHCE QDINECIPDP CHYGTCKDGI ATFTCLCRPG YTGRLCDNDI NECLSKPCLN
Dros N   PPGYTGTSCE ININDCDSNP CHRGKCIDDV NSFKCLCDPG YTGYICQKQI NECESNPCQF
```

```
CISNPCHKGA LCDTNPLNGQ YICTCPQGYK GADCTEDVDE CAMANSNPCE HAGKCVNTDG   438
CISNPCNEGS NCDTNPVNGK AICTCPSGYT GPACSQDVDE CSLG-ANPCE HAGKCINTLG   434
CISNPCNEGS NCDTNPVNGK AICTCPPGYT GPACNNDVDE CSLG-ANPCE HGGRCTNTLG   433
CTSNPCHADA ICDTSPINGS YACSCATGYK GVDCSEDIDE CDQG—SPCE HNGICVNTPG    470

QSNPCVNNGQ CVDKVNRFQC LCPPGFTGPV CQIDIDDCSS TPCLNGAKCI DHPNGYECQC   558
ASSPCLHNGR CLDKINEFQC ECPTGFTGHL CQYDVDECAS TPCKNGAKCL DGPNTYTCVC   554
ASNPCLHNGK CIDKINEFRC DCPTGFSGNL CQHDFDECTS TPCKNGAKCL DGPNSYTCQC   553
QSNPCLNDGT CHDKINGFKC SCALGFTGAR CQINIDDCQS QPCRNRGICH DSIAGYSCEC   590

DGRCIDLVNG YQCNCQPGTS GVNCEINFDD CASNPCIHG- ICMDGINRYS CVCSPGFTGQ   677
RGTCQDPDNA YLCFCLKGTT GPNCEINLDD CASSPCDSG- TCLDKIDGYE CACEPGYTGS   673
GGQCTDRENG YICTCPKGTT GVNCETKIDD CASNLCDNG- KCIDKIDGYE CTCEPGYTGK   672
DGHCQDRVGS YYCQCQAGTS GKNCEVNVNE CHSNPCNNGA TCIDGINSYK CQCVPGFTGQ   710
```

FIG.13B

45

```
hum N    RCNIDIDECA SNPCRKGATC INGVNGFRCI CPEGPHHPSC YSQVNECLSN PCI-HGNCTG
TAN-1    MCNSNIDECA GNPCHNGGTC EDGINGFTCR CPEGYHDPTC LSEVNECNSN PCV-HGACRD
Xen N    LCNININECD SNPCRNGGTC KDQINGFTCV CPDGYHDHMC LSEVNECNSN PCI-HGACHD
Dros N   HCEKNVDECI SSPCANNGVC IDQVNGYKCE CPRGFYDAHC LSDVDECASN PCVNEGRCED
```

```
hum N    DECASNPCLN QGTCFDDISG YTCHCVLPYT GKNCQTVLAP CSPNPCENAA VCKESPNFES
TAN-1    NECASNPCLN KGTCIDDVAG YKCNCLLPYT GATCEVVLAP CAPSPCRNGG ECRQSEDYES
Xen N    NECSSNPCLN HGTCIDDVAG YKCNCMLPYT GAICEAVLAP CAGSPCKNGG RCKESEDFET
Dros N   DDCVTNPCGN GGTCIDKVNG YKCVCKVPFT GRDCESKMDP CASNRCKNEA KCTPSSNFLD
```

```
hum N    CLANPCQNGG SCMDGVNTFS CLCLPGFTGD KCQTDNMECL SEPCKNGGTC SDYVNSYTCK
TAN-1    CRPNPCHNGG SCTDGINTAF CDCLPGFRGT FCEEDINECA SDPCRNGANC TDCVDSYTCT
Xen N    CQPNPCHNGG SCSDGINMFF CNCPAGFRGP KCEEDINECA SNPCKNGANC TDCVNSYTCT
Dros N   CASFPCQNGG TCLDGIGDYS CLCVDGFDGK HCETDINECL SQPCQNGATC SQYVNSYTCT
```

```
GLSGYKCLCD AGWVGINCEV DKNECLSNPC QNGGTCDNLV NGYRCTCKKG FKGYNCQVNI   796
SLNGYKCDCD PGWSGTNCDI NNNECESNPC VNGGTCKDMT SGIVCTCREG FSGPNCQTNI   792
GVNGYKCDCE AGWSGSNCDI NNNECESNPC MNGGTCKDMT GAYICTCKAG FSGPNCQTNI   791
GINEFICHCP PGYTGKRCEL DIDECSSNPC QHGGTCYDKL NAFSCQCMPG YTGQKCETNI   830
```

```
YTCLCA-PGW QGQRCTIDID EC-ISKPCMN HGLCHNTQGS YMCECPPGFS GMDCEEDIDD   914
FSCVCPTAGA KGQTCEVDIN EC-VLSPCRH GASCQNTHGG YRCHCQAGYS GRNCETDIDD   911
FSCECP-PGW QGQTCEIDMN EC-VNRPCRN GATCQNTNGS YKCNCKPGYT GRNCEMDIDD   909
FSCTCK-LGY TGRYCDEDID ECSLSSPCRN GASCLNVPGS YRCLCTKGYE GRDCAINTDD   949
```

```
CQAGFDGVHC ENNINECTES SCFNGGTCVD GINSFSCLCP VGFTGSFCLH EINECSSHPC   1034
CPAGFSGIHC ENNTPDCTES SCFNGGTCVD GINSFTCLCP PGFTGSYCQH VVNECDSRPC   1031
CQPGFSGIHC ESNTPDCTES SCFNGGTCID GINTFTCQCP PGFTGSYCQH DINECDSKPC   1029
CPLGFSGINC QTNDEDCTES SCLNGGSCID GINGYNCSCL AGYSGANCQY KLNKCDSNPC   1069
```

FIG.13C

```
hum N    LNEGTCVDGL GTYRCSCPLG YTGKNCQTLV NLCSRSPCKN KGTCVQKKAE SQCLCPSGWA
TAN-1    LLGGTCQDGR GLHRCTCPQG YTGPNCQNLV HWCDSSPCKN GGKCWQTHTQ YRCECPSGWT
Xen N    LNGGTCQDSY GTYKCTCPQG YTGLNCQNLV RWCDSSPCKN GGKCWQTNNF YRCECKSGWT
Dros N   LNGATCHEQN NEYTCHCPSG FTGKQCSEYV DWCGQSPCEN GATCSQMKHQ FSCKCSAGWT

hum N    SNPCQHGATC SDFIGGYRCE CVPGYQGVNC EYEVDECQNQ PCQNGGTCID LVNHFKCSCP
TAN-1    PSPCQNGATC TDYLGGYSCK CVAGYHGVNC SEEIDECLSH PCQNGGTCLD LPNTYKCSCP
Xen N    PNPCQNGATC TDYLGGYSCE CVAGYHGVNC SEEINECLSH PCQNGGTCID LINTYKCSCP
Dros N   SQPCQNGGTC RDLIGAYECQ CRQGFQGQNC ELNIDDCAPN PCQNGGTCHD RVMNFSCSCP

hum N    CLSNPCSSEG SLDCIQLTND YLCVCRSAFT GRHCETFVDV CPQMPCLNGG TCAVASNMPD
TAN-1    CLSNPCDARG TQNCVQRVND FHCECRAGHT GRRCESVING CKGKPCKNGG TCAVASNTAR
Xen N    CLSNPCDSRG TQNCIQLVND YRCECRQGFT GRRCESVVDG CKGMPCRNGG TCAVASNTER
Dros N   CLSNPCSNAG TLDCVQLVNN YHCNCRPGHM GRHCEHKVDF CAQSPCQNGG NCNI——RQS
```

```
GAYCDVPNVS CDIAASRRGV LVEHLCQHSG VCINAGNTHY CQCPLGYTGS YCEEQLDECA    1154
GLYCDVPSVS CEVAAQRQGV DVARLCQHGG LCVDAGNTHH CRCQAGYTGS YCEDLVDECS    1151
GVYCDVPSVS CEVAAKQQGV DIVHLCRNSG MCVDTGNTHF CRCQAGYTGS YCEEQVDECS    1149
GKLCDVQTIS CQDAADRKGL SLRQLC-NNG TCKDYGNSHV CYCSQGYAGS YCQKEIDECQ    1188

PGTRGLLCEE NIDDCAR——  ——GPHCLN GGQCMDRIGG YSCRCLPGFA GERCEGDINE    1267
RGTQGVHCEI NVDDCNPPVD PVSRSPKCFN NGTCVDQVGG YSCTCPPGFV GERCEGDVNE    1271
RGTQGVHCEI NVDDCTPFYD SFTLEPKCFN NGKCIDRVGG YNCICPPGFV GERCEGDVNE    1269
PGTMGIICEI NKDDCKP——  ——GACHN NGSCIDRVGG FECVCQPGFV GARCEGDINE    1300

GFICRCPPGF SGARCQS——  SCGQVKCRKG EQCVHTAS— GPRCFCPSP- —RDCES——    1376
GFICKCPAGF EGATCENDAR TCGSLRCLNG GTCISGPR— SPTCLCLGPF TGPECQFPAS    1389
GFICKCPPGF DGATCEYDSR TCSNLRCQNG GTCISVLT— SSKCVCSEGY TGATCQYPVI    1387
GHHCICNNGF YGKNCELSGQ DCDSNPCRVG -NCVVADEGF GYRCECPRGT LGEHCEIDTL    1415
```

FIG.13D

47

```
hum N    -GC-ASSPCQ HGGSCHPQRQ PPYYSCQCAP PFSGSRCEL  -YTAPP—— -S——TPP
TAN-1    SPCLGGNPCY NQGTCEPTSE SPFYRCLCPA KFNGLLCHIL DYSFGG—— -GAGRDIPPP
Xen N    SPC-ASHPCY NGGTCQFFAE EPFFQCFCPK NFNGLFCHIL DYEFPG—— -GLGKNITPP
Dros N   DEC-SPNPCA QGAACEDLLG D—YECLCPS KWKGKRCDIY DANYPGWNGG SGSGNDRYAA

hum N    NN-QCDELCN TVECLFDNFE CQGNSKTCK- -YDKYCADHF KDNHCNQGCN SEECGWDGLD
TAN-1    SDGHCDSQCN SAGCLFDGFD CQRAEGQCNP LYDQYCKDHF SDGHCDQGCN SAECEWDGLD
Xen N    NDGKCDSQCN NTGCLYDGFD CQKVEVQCNP LYDQYCKDHF QDGHCDQGCN NAECEWDGLD
Dros N   KNGKCNEECN NAACHYDGHD CERKLKSCDS LFDAYCQKHY GDGFCDYGCN NAECSWDGLD

hum N    YYGEKSAAMK KQ—R—— —————— ——MTRRSL PGEQ——E QEVAGSKVFL
TAN-1    YYGREEELRK HPIKRAAEGW AAPDALLGQV KASLLPGGSE GGRRRRELDP MDVRGSIVYL
Xen N    YYGNEEELKK HHIKRSTDYW SDAPSAI— -FSTMKESIL LGRHRRELDE MEVRGSIVYL
Dros N   WKDNVRVPEI EDTDFARKNK ILYTQQVHQ— —————— ————— —TGIQIYL
```

LNR (Notch/Lin-12 Repeats)

```
—A—TCL SQYCADKARD GVCDEACNSH ACQWDGGDCS LTMENPWANC SSPLPCWDYI  1476
LIEE—ACE LPECQEDAGN KVCSLQCNNH ACGWDGGDCS LNFNDPWKNC TQSLQCWKYF  1501
DNDD—ICE NEQCSELADN KVCNANCNNH ACGWDGGDCS LNFNDPWKNC TQSLQCWKYF  1498
DLEQQRAMCD KRGCTEKQGN GICDSDCNTY ACNFDGNDCS LGI-NPWANC TAN-EXWNKF  1531

CAADQPEN-L AEGTLVIVVL MPPEQLLQDA R-SFLRALGT LLHTNLRIKR DSQGELMVYP  1591
CAEHVPER-L AAGTL-VVVV LMPPEQLRNS SFHFLRELSR VLHTNVVFKR DAHGQQMIFP  1619
C-ANMPEN-L AEGTLVLVVL MPPERLKNNS V-NFLRELSR VLHTNVVFKK DSKGEYKIYP  1615
CENKTQSPVL AEGAMSVVML MNVEAFREIQ A-QFLRNMSH MLRTTVRLKK DALGHDIIIN  1650

                                                              TM
EIDNRQCVQD SDHCFKNTDA AAALLASHAI QG——TLSYP LVSVVSESLT PERT-Q-LLY  1680
EIDNRQCVQA SSQCFQSATD VAAFLGALAS LGSL-NIPYK IEAVQSETVE PPPPAQ-LHF  1737
EIDNRQCYKS SSQCFNSATD VAAFLGALAS LGSLDTLSYK IEAVKSENME TPKPST-LYP  1730
EIDNRKCTEC FTHAVEAAEF LAATAAKHQL RNDFQ-IHSV RGIKNPGDED NGEPPANVKY  1745
```

FIG.13E

48

hum N	LLAVAVVIIL FIILLGVIMA	KRKRK—HGS	LWLPEGFTLR	RDASNHKRRE	PVGQDAVGLK
TAN-1	MYVAAAAFVL LFFVGCGVLL	SRKRRRQHGQ	LWFPEGFKV-	SEASKKKRRE	ELGEDSVGLK
Xen N	MLSMLVIPLL IIFVFMMVIV	NKKRRREHDS	FGSPTALFQK	NPA-KRNGET	PW-EDSVGLK
Dros N	VITGIILVII ALAFFGMVL-	STQRKRAHGV	TWFPEGFRAP	AAVMSRRRRD	PHGQEMRNLN

CDC-10/Ankyrin Repeats

hum N	PIDRRPWTQQ	HLEAADIRRT	PSLALTPPQA	EQEVDVLDVN	VRGPDGCTPL MLASLRGGSS
TAN-1	QTDHRQWTQQ	HLDAADL-RM	SAMAPTPPQG	EVDADCMDVN	VRGPDGFTPL MIASCSGGGL
Xen N	KTDPRQWTRQ	HLDAADL-RI	SSMAPTPPQG	EIEADCMDVN	VRGPDGFTPL MIASCSGGGL
Dros N	EADQRVWSQA	HLDVVDV-R-	AIM—TPP-A	HQDGGKHDVD	ARGPCGLTPL MIAAVRGGGL

hum N	ANAQDNMGRC	PLHAAVAADA	QGVFQILIRN	RVTDLDARMN	DGTTPLILAA RLAVEGMVAE
TAN-1	ANIQDNMGRT	PLHAAVSADA	QGVFQILIRN	RATDLDARMH	DGTTPLILAA RLAVEGMLED
Xen N	ANVQDNMGRT	PLHAAVAADA	QGVFQILIRN	RATDLDARMF	DGTTPLILAA RLAVEGMVEE
Dros N	ANCQDNTGRT	PLHAAVAADA	MGVFQILLRN	RATNLNARMH	DGTTPLILAA RLAIEGMVED

NLSVQVSEAN	LIGTGTSEHW	VDDE———	———G	PQPKKVKAED	EALLSE-EDD	1782
PLK-NASDGA	LMDDNQNE-W	GDED———	————	LETKKFRFEE	PVVLPD-LDD	1837
PIK-NMTDGS	FMDDNQNE-W	GDEET———	————	LENKRFRFEE	QVILPELVDD	1831
KQVAMQSQGV	GQPGAH——W	SDDESDMPLP	KRQRSDPVSG	VGLGNNGGYA	SDHTMVSEYE	1861

DLSDEDEDAE	DSSANIITDL	VYQGASLQAQ	TDRTGEMALH	LAARYSRADA	AKRLLDAGAD	1902
ETGNSEEE-E	DAPA-VISDF	IYQGASLHNQ	TDRTGETALH	LAARYSRSDA	AKRLLEASAD	1954
ETGNSEEE-E	DASANMISDF	IGQGAQLHNQ	TDRTGETALH	LAARYARADA	AKRLLESSAD	1949
DTGEDIENNE	DSTAQVISDL	LAQGAELNAT	MDKTGETSLH	LAARFARADA	AKRLLDAGAD	1976

LINCQADVNA	VDDHGKSALH	WAAAVNNVEA	TLLLLKNGAN	RDMQDNKEET	PLFLAAREGS	2022
LINSHADVNA	VDDLGKSALH	WAAAVNNVDA	AVVLLKNGAN	KDMQNNREET	PLFLAAREGS	2074
LINAHADVNA	VDEFGKSALH	WAAAVNNVDA	AAVLLKNSAN	KDMQNNKEET	SLFLAAREGS	2069
LITADADINA	ADNSGKTALH	WAAAVNNTEA	VNILLMHHAN	RDAQDDKDET	PLFLAAREGS	2096

FIG.13F

49

```
hum N   YEAAKILLDH FANRDITDHM DRLPRDVARD RMHHDIVRLL DEYNVTPSPP —GTVL—TS
TAN-1   YETAKVLLDH FANRDITDHM DRLPRDIAQE RMHHDIVRLL DEYNLVRSPQ LHGAPLGGTP
Xen N   YETAKVLLDH YANRDITDHM DRLPRDIAQE RMHHDIVHLL DEYNLVKSPT LHNGPLGAT-
Dros N  YEACKALLDN FANREITDHM DRLPRDVASE RLHHDIVRLL DE-HVPRSPQ MLSMTPQAMI

             NLS                    CK II   cdc2      cdc2
hum N   GSRRKKSLSE KVQLSE—SS VTLSPVDSLE SPHTYVSDTT SSPM ——————— ———————
TAN-1   A-RRKKSQDG KGCLLD—SS GMLSPVDSLE SPHGYLSDVA SPPL ——————— ———————
Xen N   A-RRKKSQDG KTTLLDSGSS GVLSPVDSLE STHGYLSDVS SPPL ——————— ———————
Dros N  GS-PDNGLDA TGSLRRKASS KKTSAASKKA ANLNGLNPGQ LTGGVSGVPG VPPTNSAAQA
        BNTS

hum N   ——————— ——————— ——————— ITSPGILQAS PNPML—ATA APPAPVHAQH
TAN-1   ——————— ——————— ——————— LPSPF—QQS PSVPLNHLPG MPDTHLGIGH
Xen N   ——————— ——————— ——————— MTSPF—QQS PSMPLNHLTS MPESQLGMNH
Dros N  YEDCIKNAQS MQSLQGNGLD MIKLDNYAYS MGSPF—QQE LLNGQGLGMN GNGQRNGVGP
        CK II                              cdc2
```

```
ALSPV——— ———————ICGP NRSFLSLKHT PMGKKSRRPS AKSTMPTSLP NLAKEAKDAK   2127
TLSPP——— ———————LCSP NGYLGSLKPG VQGKKVRKPS SKGLACGS— —KEAKDLK     2178
TLSPP——— ———————ICSP NGYMGNMKPS VQSKKARKPS IKGNGC—— —KEAKELK    2170
GSPPPGQQQP QLITQPTVIS AGNGGNNGNG NASGKQSNQT AKQKAA—— —KKAKLIE    2208

——————— ——————— ——————— ——————— ——————— ——————— 2169
——————— ——————— ——————— ——————— ——————— ——————— 2219
——————— ——————— ——————— ——————— ——————— ——————— 2213
AAAAAAAVAA MSHELEGSPV GVGMGGNLPS PYDTSSMYSN AMAAPLANGN PNTGAKQPPS   2327

ALSFSNLHEM Q——— ——————— — -PLAHGASTV LPSVSQLLSH HHIVSPCS—   2235
LNVAA-KPEM AALGGGGRLA FETGPPRLSH LPVASGTSTV LGSSSGGALN FTVGGSTSLN   2306
INMAT-KQEM AA—GSNRMA FDAMVPRLTH L-NASSPNTI MS——NGSMH FTVGGAPTMN  2294
GVLPGGLCGM GGLSGAGNGN SHEQGLSPPY SNQSPPHSVQ SSLALSPHAY LGSPSPAKSR   2445
```

FIG.13G

```
hum N    GSAGSLSRLH PVPVPADW—  MNRMEVNETQ YNEMFGMVLA PAEG-THPGI APQSRPPEGK
TAN-1    GQCEWLSRLQ SGMVPNQYNP LRGSVAPGPL STQAPSLQHG -MVGPLHSSL AASALSQMMS
Xen N    SQCDWLARLQ NGMVQNQYDP IRNGIQQGN- AQQAQALQHG LMTS-LHNGL PATTLSQMMT
Dros N   PSLPTSPTHI QAMRHATQQK QFGGSNLNSL LGGANGGGVV GGGGGGGGGV GQGPQNSPVS


hum N    APQPQSTCPP AVAGPLPTMY QIP——————EM ARL-PSVAFP TAMMPQQDGQ VAQTILPAYH
TAN-1    PPQPHLGVSS AASGHLGRSF LSGEPSQADV QPLGPSSLAV HTILPQ-ESP ALPTSLPSSL
Xen N    MQQQHHN-SS TTSTHINSPF CSSDISQTDL QQM——SSNNI HSVMPQ-DTQ IFAASLPSNL
Dros N   QQQLGGLEFG SAGLDLNG-F CGSPDSFHSG QMNPPS——I QSSMSG-SSP STNMLSPSSQ


hum N    SDWSDVTTSP TPGGAGGGQR GPGTHMSEPPHNN MQVYA
TAN-1    SDWSEGVSSP PT———SMQ SQIARIPEAFK
Xen N    SDWSEGISSP PT———SMQ PQRTHIPEAFK
Dros N   SDWSEGVQSP AANNLYISGG HQANKGSEAIYI
```

```
         ————————   ——HITTPRE PLPP-IV-TF QLIPKGSIAQ PAG————————   ———————————   2320
         ————————   -YQGLPSTRL ATQPHLVQTQ QVQPQNLQMQ QQNLQPANIQ QQQSLQPPPP   2414
         ————————   -YQAMPNTRL ANQPHLMQAQ QMQQQQN——   ———————————   ———LQLHQS   2384
LGIISPTGSD MGIMLAPPQS SKNSAIMQTI SPQQQQQQQQ QQQQQHQQQQ QQQQQQQQQQ   2565
```

PEST -containing Region

```
PFPASVGKYP TPPSQHSYAS SNAAERTPSH SGHLQGEHPY LTPSPESPDQ WSSSSPHSA-   2433
VPPVTAAQFL TPPSQHSY-S S-PVENTPSH QLQVP-EGPF LTPSPESPDQ WSSSSPHSNV   2530
TQSMTTAQFL TPPSQHSY-S S-PMDNTPSH QLQVP-DHPF LTPSPESPDQ WSSSSPHSNM   2497
HNQQAFYQYL TPSSQHS——   ——GGHTPQH LVQTL-D-SY PTPSPESPGH WSSSSPRSN-   2671
```

```
                                                                 2471
                                                                 2556
                                                                 2523
                                                                 2703
```

FIG.13H

FIG.14

52
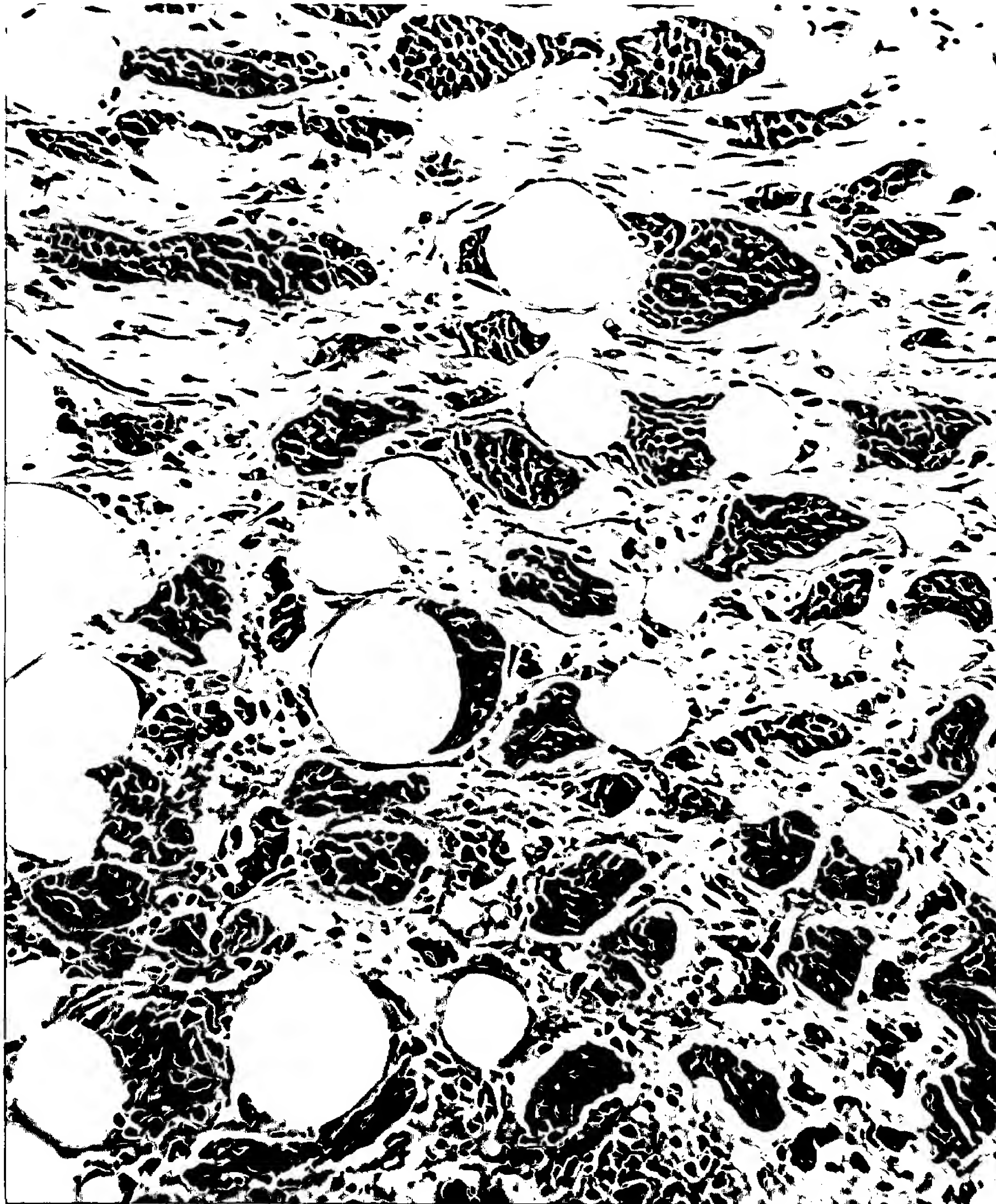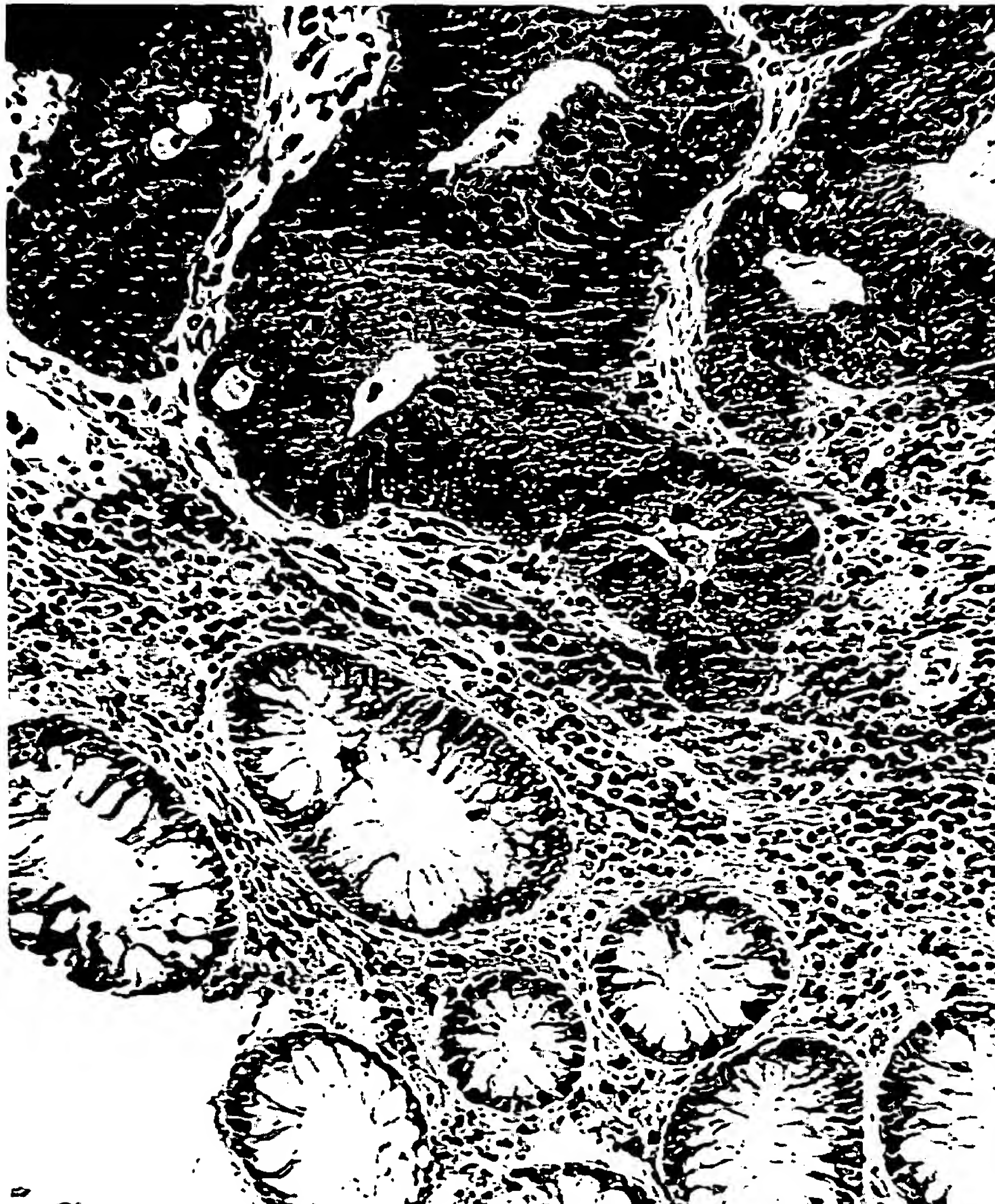
FIG.15A

FIG.15B

FIG.16A

55

FIG.16B

GGAATTCCGC CCGCCCTGCG CCCCGCTCTG CTGTGGGCGC TGCTGGCGCT CTGGCTGTGC TGCGCGGCCC CCGCCCATGC ATTGCAGTGT
  P A L R   P A L   L W A L   L A L   W L C   C A A   P A H A   L Q C>

CCAGATGGCT ATGAACCCTG TGTAAATGAA GGAATGTGTG TTACCTACCA CAATGGCACA GGATACTGCA AATGTCCAGA AGGCTTCTTG
  R D G   Y E P C   V N E   G M C   V T Y   H N G T   G Y C   K C P E   G F L>

GGGGAATATT GTCAACATCG AGACCCCTGT GAGAAGAACC GCTGCCAGAA TGGTGGGACT TGTGTGGCCC AGGCCATGCT GGGGAAAGCC
  G E Y   C Q H R   D P C   E K N   R C Q N   G G T   C V A   Q A M L   G K A>

ACGTGCCGAT GTGCCTCAGG GTTTACAGGA GAGGACTGCC AGTACTCAAC ATCTCATCCA TGCTTTGTGT CTCGACCCTG CCTGAATGGC
  T C R   C A S G   F T G   E D C   Q Y S T   S H P   C F V   S R P   C L N G>

GGCACATGCC ATATGCTCAG CCGGGATACC TATGAGTGCA CCTGTCAAGT CGGGTTTACA GGTAAGGAGT GCCAATGGAC GGATGCCTGC
  G T C   H M L S   R D T   Y E C   T C Q V   G F T   G K E   C Q W T   D A C>

CTGTCTCATC CCTGTGCAAA TGGAAGTACC TGTACCACTG TGGCCAACCA GTTCTCCTGC AAATGCCTCA CAGGCTTCAC AGGGCAGAAA
  L S H   P C A N   G S T   C T T   V A N Q   F S C   K C L   T G F T   G Q K>

TGTGAGACTG ATGTCAATGA GTGTGACATT CCAGGACACT GCCAGCATGG TGGCACCTGC CTCAACCTGC CTGGTTCCTA CCAGTGCCAG
  C E T   D V N E   C D I   P G H   C Q H G   G T C   L N L   P G S Y   Q C Q>

TGCCCTCAGG GCTTCACAGG CCAGTACTGT GACAGCCTGT ATGTGCCCTG TGCACCCTCA CCTTGTGTCA ATGGAGGCAC CTGTCGGCAG
  C P Q   G F T G   Q Y C   D S L   Y V P   C A P S   P C V   N G G T   C R Q>

ACTGGTGACT TCACTTTTGA GTGCAACTGC CTTCCAGGTT TTGAAGGGAG CACCTGTGAG AGGAATATTG ATGACTGCCC TAACCACAGG
  T G D   F T F E   C N C   L P G   F E G S   T C E   R N I   D D C   P N H R>

FIG.17A

```
        820         830         840         850         860         870         880         890         900
         •           •           •           •           •           •           •           •           •
TGTCAGAATG GAGGGGTTTG TGTGGATGGG GTCAACACTT ACAACTGCCG CTGTCCCCCA CAATGGACAG GACAGTTCTG CACAGAGGAT
  C  Q  N    G  G  V    C  V  D  G   V  N  T    Y  N  C  R    C  P  P    Q  W  T    G  Q  F  C    T  E  D>

        910         920         930         940         950         960         970         980         990
         •           •           •           •           •           •           •           •           •
GTGGATGAAT GCCTGCTGCA GCCCAATGCC TGTCAAAATG GGGGCACCTG TGCCAACCGC AATGGAGGCT ATGGCTGTGT ATGTGTCAAC
  V  D  E    C  L  L  Q   P  N  A    C  Q  N    G  G  T    C  A  N  R    N  G  G    Y  G  C  V    C  V  N>

       1000        1010        1020        1030        1040        1050        1060        1070        1080
         •           •           •           •           •           •           •           •           •
GGCTGGAGTG GAGATGACTG CAGTGAGAAC ATTGATGATT GTGCCTTCGC CTCCTGTACT CCAGGCTCCA CCTGCATCGA CCGTGTGGCC
  G  W  S    G  D  D  C   S  E  N    I  D  D    C  A  F  A    S  C  T    P  G  S    T  C  I  D    R  V  A>

        1090        1100        1110        1120        1130        1140        1150        1160        1170
         •           •           •           •           •           •           •           •           •
TCCTTCTCTT GCATGTGCCC AGAGGGGAAG GCAGGTCTCC TGTGTCATCT GGATGATGCA TGCATCAGCA ATCCTTGCCA CAAGGGGGCA
  S  F  S    C  M  C  P   E  G  K    A  G  L    L  C  H  L    D  D  A    C  I  S    N  P  C  H    K  G  A>

       1180        1190        1200        1210        1220        1230        1240        1250        1260
         •           •           •           •           •           •           •           •           •
CTGTGTGACA CCAACCCCCT AAATGGGCAA TATATTTGCA CCTGCCCACA AGGCTACAAA GGGGCTGACT GCACAGAAGA TGTGGATGAA
  L  C  D    T  N  P  L   N  G  Q    Y  I  C    T  C  P  Q    G  Y  K    G  A  D    C  T  E  D    V  D  E>

       1270        1280        1290        1300        1310        1320        1330        1340        1350
         •           •           •           •           •           •           •           •           •
TGTGCCATGG CCAATAGCAA TCCTTGTGAG CATGCAGGAA AATGTGTGAA CACGGATGGC GCCTTCCACT GTGAGTGTCT GAAGGGTTAT
  C  A  M    A  N  S  N   P  C  E    H  A  G    K  C  V  N    T  D  G    A  F  H    C  E  C  L    K  G  Y>

       1360        1370        1380        1390        1400        1410        1420        1430        1440
         •           •           •           •           •           •           •           •           •
GCAGGACCTC GTTGTGAGAT GGACATCAAT GAGTGCCATT CAGACCCCTG CCAGAATGAT GCTACCTGTC TGGATAAGAT TGGAGGCTTC
  A  G  P    R  C  E  M   D  I  N    E  C  H    S  D  P  C    Q  N  D    A  T  C    L  D  K  I    G  G  F>

       1450        1460        1470        1480        1490        1500        1510        1520        1530
         •           •           •           •           •           •           •           •           •
ACATGTCTGT GCATGCCAGG TTTCAAAGGT GTGCATTGTG AATTAGAAAT AAATGAATGT CAGAGCAACC CTTGTGTGAA CAATGGGCAG
  T  C  L    C  M  P  G   F  K  G    V  H  C    E  L  E  I    N  E  C    Q  S  N    P  C  V  N    N  G  Q>

       1540        1550        1560        1570        1580        1590        1600        1610        1620
         •           •           •           •           •           •           •           •           •
TGTGTGGATA AAGTCAATCG TTTCCAGTGC CTGTGTCCTC CTGGTTTCAC TGGGCCAGTT TGCCAGATTG ATATTGATGA CTGTTCCAGT
  C  V  D    K  V  N  R   F  Q  C    L  C  P    P  G  F  T    G  P  V    C  Q  I    D  I  D  D    C  S  S>
```

FIG.17B

```
        1630      1640      1650      1660      1670      1680      1690      1700      1710
         •         •         •         •         •         •         •         •         •
ACTCCCTGTC TGAATGGGGC AAAGTGTATC GATCACCCGA ATGGCTATGA ATGCCAGTGT GCCACAGGTT TCACTGGTGT GTTGTGTGAC
 T P C     L N G A     K C I     D H P     N G Y E     C Q C     A T G     F T G V     L C E>

        1720      1730      1740      1750      1760      1770      1780      1790      1800
         •         •         •         •         •         •         •         •         •
GAGAACATTG ACAACTGTGA CCCCGATCCT TGCCACCATG GTCAGTGTCA GGATGGTATT GATTCCTACA CCTGCATCTG CAATCCCGGG
 E N I     D N C D     P D P     C H H     G Q C Q     D G I     D S Y     T C I     C N P G>

        1810      1820      1830      1840      1850      1860      1870      1880      1890
         •         •         •         •         •         •         •         •         •
TACATGGGCG CCATCTGCAG TGACCAGATT GATGAATGTT ACAGCAGCCC TTGCCTGAAC GATGGTCGCT GCATTGACCT GGTCAATGGC
 Y M G     A I C S     D Q I     D E C     Y S S P     C L N     D G R     C I D L     V N G>

        1900      1910      1920      1930      1940      1950      1960      1970      1980
         •         •         •         •         •         •         •         •         •
TACCAGTGCA ACTGCCAGCC AGGCACGTCA GGGGTTAATT GTGAAATTAA TTTTGATGAC TGTGCAAGTA ACCCTTGTAT CCATGGAATC
 Y Q C     N C Q P     G T S     G V N     C E I N     F D D     C A S     N P C I     H G I>

        1990      2000      2010      2020      2030      2040      2050      2060      2070
         •         •         •         •         •         •         •         •         •
TGTATGGATG GCATTAATCG CTACAGTTGT GTCTGCTCAC CAGGATTCAC AGGGCAGAGA TGTAACATTG ACATTGATGA GTGTGCCTCC
 C M D     G I N R     Y S C     V C S     P G F T     G Q R     C N I     D I D E     C A S>

        2080      2090      2100      2110      2120      2130      2140      2150      2160
         •         •         •         •         •         •         •         •         •
AATCCCTGTC GCAAGGGTGC AACATGTATC AACGGTGTGA ATGGTTTCCG CTGTATATGC CCCGAGGGAC CCCATCACCC CAGCTGCTAC
 N P C     R K G A     T C I     N G V     N G F R     C I C     P E G     P H H P     S C Y>

        2170      2180      2190      2200      2210      2220      2230      2240      2250
         •         •         •         •         •         •         •         •         •
TCACAGGTGA ACGAATGCCT GAGCAATCCC TGCATCCATG GAAACTGTAC TGGAGGTCTC AGTGGATATA AGTGTCTCTG TGATGCAGGC
 S Q V     N E C L     S N P     C I H     G N C T     G G L     S G Y     K C L C     D A G>

        2260      2270      2280      2290      2300      2310      2320      2330      2340
         •         •         •         •         •         •         •         •         •
TGGCTTGGCA TCAACTGTGA ACTGGACAAA AATGAATGCC TTTCGAATCC ATGCCAGAAT GGAGGAACTT GTGACAATCT GGTGAATGGA
 W V G     I N C E     V D K     N E C     L S N P     C Q N     G G T     C D N L     V N G>

        2350      2360      2370      2380      2390      2400      2410      2420      2430
         •         •         •         •         •         •         •         •         •
TACAGGTGTA CTTGCAAGAA GGGCTTTAAA GGCTATAACT GCCAGGTGAA TATTGATGAA TGTGCCTCAA ATCCATGCCT GAACCAAGGA
 Y R C     T C K F     G F K     G Y N     C Q V N     I D E     C A S     N P C L     N Q G>
```

FIG.17C

```
      2440       2450       2460       2470       2480       2490       2500       2510       2520
ACCTGCTTTG ATGACATAAG TGGCTACACT TGCCACTGTG TGCTGCCATA CACAGGCAAG AATTGTCAGA CAGTATTGGC TCCCTGTTCC
 T  C  F    D  D  I  S    G  Y  T    C  H  C    V  L  P  Y    T  G  K    N  C  Q    T  V  L  A    P  C  S>

      2530       2540       2550       2560       2570       2580       2590       2600       2610
CCAAACCCTT GTGAGAATGC TGCTGTTTGC AAAGAGTCAC CAAATTTTGA GAGTTATACT TGCTTGTGTG CTCCTGGCTG GCAAGGTCAG
 P  N  P    C  E  N  A    A  V  C    K  E  S    P  N  F  E    S  Y  T    C  L  C    A  P  G  W    Q  G  Q>

      2620       2630       2640       2650       2660       2670       2680       2690       2700
CGGTGTACCA TTGACATTGA CGAGTGTATC TCCAAGCCCT GCATGAACCA TGGTCTCTGC CATAACACCC AGGGCAGCTA CATGTGTGAA
 R  C  T    I  D  I  D    E  C  I    S  K  P    C  M  N  H    G  L  C    H  N  T    Q  G  S  Y    M  C  E>

      2710       2720       2730       2740       2750       2760       2770       2780       2790
TGTCCACCAG GCTTCAGTGG TATGGACTGT GAGGAGGACA TTGATGACTG CCTTGCCAAT CCTTGCCAGA ATGGAGGTTC CTGTATGGAT
 C  P  P    G  F  S  G    M  D  C    E  E  D    I  D  D  C    L  A  N    P  C  Q    N  G  G  S    C  M  D>

      2800       2810       2820       2830       2840       2850       2860       2870       2880
GGAGTGAATA CTTTCTCCTG CCTCTGCCTT CCGGGTTTCA CTGGGGATAA GTGCCAGACA GACATGAATG AGTGTCTGAG TGAACCCTGT
 G  V  N    T  F  S  C    L  C  L    P  G  F    T  G  D  K    C  Q  T    D  M  N    E  C  L  S    E  P  C>

      2890       2900       2910       2920       2930       2940       2950       2960       2970
AAGAATGGAG GGACCTGCTC TGACTACGTC AACAGTTACA CTTGCAAGTG CCAGGCAGGA TTTGATGGAG TCCATTGTGA GAACAACATC
 K  N  G    G  T  C  S    D  Y  V    N  S  Y    T  C  K  C    Q  A  G    F  D  G    V  H  C  E    N  N  I>

      2980       2990       3000       3010       3020       3030       3040       3050       3060
AATGAGTGCA CTGAGAGCTC CTGTTTCAAT GGTGGCACAT GTGTTGATGG GATTAACTCC TTCTCTTGCT TGTGCCCTGT GGGTTTCACT
 N  E  C    T  E  S  S    C  F  N    G  G  T    C  V  D  G    I  N  S    F  S  C    L  C  P  V    G  F  T>

      3070       3080       3090       3100       3110       3120       3130       3140       3150
GGATCCTTCT GCCTCCATGA GATCAATGAA TGCAGCTCTC ATCCATGCCT GAATGAGGGA ACGTGTGTTG ATGGCCTGGG TACCTACCGC
 G  S  F    C  L  H  E    I  N  E    C  S  S    H  P  C  L    N  E  G    T  C  V    D  G  L  G    T  Y  R>

      3160       3170       3180       3190       3200       3210       3220       3230       3240
TGCAGCTGCC CCCTGGGCTA CACTGGGAAA AACTGTCAGA CCCTGGTGAA TCTCTGCAGT CGGTCTCCAT GTAAAAACAA AGGTACTTGT
 C  S  C    P  L  G  Y    T  G  K    N  C  Q    T  L  V  N    L  C  S    R  S  P    C  K  N  K    G  T  C>
```

FIG.17D

```
         3250       3260       3270       3280       3290       3300       3310       3320       3330
          •          •          •          •          •          •          •          •          •
GTTCAGAAAA AAGCAGAGTC CCAGTGCCTA TGTCCATCTG GATGGGCTGG TGCCTATTGT GACGTGCCCA ATGTCTCTTG TGACATAGCA
 V  Q  K    K  A  E  S    Q  C  L    C  P  S   G  W  A   G  A  Y  C   D  V  P    N  V  S   C  D  I  A>

         3340       3350       3360       3370       3380       3390       3400       3410       3420
          •          •          •          •          •          •          •          •          •
GCCTCCAGGA GAGGTGTGCT TGTTGAACAC TTGTGCCAGC ACTCACGTGT CTGCATCAAT GCTGGCAACA CGCATTACTG TCAGTGCCCC
 A  S  R    R  G  V  L    V  E  H   L  C  Q    H  S  G   V  C  I  N   A  G  N    T  H  Y  C  Q  C  P>

         3430       3440       3450       3460       3470       3480       3490       3500       3510
          •          •          •          •          •          •          •          •          •
CTGGGCTATA CTGGGAGCTA CTGTGAGGAG CAACTCGATG AGTGTGCGTC CAACCCCTGC CAGCACGGGG CAACATGCAG TGACTTCATT
 L  G  Y    T  G  S  Y    C  E  E    Q  L  D   E  C  A  S   N  P  C    Q  H  G   A  T  C  S  D  F  I>

         3520       3530       3540       3550       3560       3570       3580       3590       3600
          •          •          •          •          •          •          •          •          •
GGTGGATACA GATGCCAGTG TGTCCCAGGC TATCAGGGTG TCAACTGTGA GTATGAAGTG GATGAGTGCC AGAATCAGCC CTGCCAGAAT
 G  G  Y    R  C  Q  C    V  P  G   Y  Q  G    V  N  C   E  Y  E  V   D  E  C    Q  N  Q  P  C  Q  N>

         3610       3620       3630       3640       3650       3660       3670       3680       3690
          •          •          •          •          •          •          •          •          •
GGAGGCACCT GTATTGACCT TGTGAACCAT TTCAAGTGCT CTTGCCCACC AGGCACTCGG GGCCTACTCT GTGAAGAGAA CATTGATGAC
 G  G  T    C  I  D  L    V  N  H   F  K  C    S  C  P  P   G  T  R    G  L  L   C  E  E  N  I  D  D>

         3700       3710       3720       3730       3740       3750       3760       3770       3780
          •          •          •          •          •          •          •          •          •
TGTCCCCGGG GTCCCCATTG CCTTAATGGT GGTCAGTGCA TGGATAGGAT TGGAGGCTAC AGTTGTCGCT GCTTGCCTGG CTTTGCTGGG
 C  P  R    G  P  H  C    L  N  G   G  Q  C    M  D  R  I   G  G  Y    S  C  R   C  L  P  G  F  A  G>

         3790       3800       3810       3820       3830       3840       3850       3860       3870
          •          •          •          •          •          •          •          •          •
GAGCGTTGTG AGGGAGACAT CAACGAGTGC CTCTCCAACC CCTGCAGCTC TGAGGGCAGC CTGGACTGTA TACAGCTCAC CAATGACTAC
 E  R  C    E  G  D  I    N  E  C   L  S  N    P  C  S  S   E  G  S    L  D  C   I  Q  L  T  N  D  Y>

         3880       3890       3900       3910       3920       3930       3940       3950       3960
          •          •          •          •          •          •          •          •          •
CTGTGTGTTT GCCGTAGTGC CTTTACTGGC CGGCACTGTG AAACCTTCGT CGATGTGTGT CCCCAGATGC CCTGCCTGAA TGGAGGGACT
 L  C  V    C  R  S  A    F  T  G   R  H  C    E  T  F  V   D  V  C    P  Q  M   P  C  L  N  G  G  T>

         3970       3980       3990       4000       4010       4020       4030       4040       4050
          •          •          •          •          •          •          •          •          •
TGTGCTGTGG CCAGTAACAT GCCTGATGGT TTCATTTGCC GTTGTCCCCC GGGATTTTCC GGGGCAAGGT GCCAGAGCAG CTGTGGACAA
 C  A  V    A  S  N  M    P  D  G   F  I  C    R  C  P  P   G  F  S    G  A  R   C  Q  S  S  C  G  Q>
```

FIG.17E

61

```
           4060        4070        4080        4090        4100        4110        4120        4130        4140
            •           •           •           •           •           •           •           •           •
GTGAAATGTA GGAAGGGGGA GCCAGTGTGTG CACACCGCCT CTGGACCCCG CTGCTTCTGC CCCAGTCCCC GGGACTGCGA GTCAGGCTGT
 V  K  C    R  K  G  E    Q  C  V    H  T  A    S  G  P  R    C  F  C    P  S  P    R  D  C  E    S  G  C>

           4150        4160        4170        4180        4190        4200        4210        4220        4230
            •           •           •           •           •           •           •           •           •
GCCAGTAGCC CCTGCCAGCA CGGGGGCAGC TGCCACCCTC AGCGCCAGCC TCCTTATTAC TCCTGCCAGT GTGCCCCACC ATTCTCGGGT
 A  S  S    P  C  Q  H    G  G  S    C  H  P    Q  R  Q  P    P  Y  Y    S  C  Q    C  A  P  P    F  S  G>

           4240        4250        4260        4270        4280        4290        4300        4310        4320
            •           •           •           •           •           •           •           •           •
AGCCGCTGTG AACTCTACAC GGCACCCCCC AGCACCCCTC CTGCCACCTG TCTGAGCCAG TATTGTGCCG ACAAAGCTCG GGATGGCGTC
 S  R  C    E  L  Y  T    A  P  P    S  T  P    P  A  T  C    L  S  Q    Y  C  A    D  K  A  R    D  G  V>

           4330        4340        4350        4360        4370        4380        4390        4400        4410
            •           •           •           •           •           •           •           •           •
TGTGATGAGG CCTGCAACAG CCATGCCTGC CAGTGGGATG GGGGTGACTG TTCTCTCACC ATGGAGAACC CCTGGGCCAA CTGCTCCTCC
 C  D  E    A  C  N  S    H  A  C    Q  W  D    G  G  D  C    S  L  T    M  E  N    P  W  A  N    C  S  S>

           4420        4430        4440        4450        4460        4470        4480        4490        4500
            •           •           •           •           •           •           •           •           •
CCACTTCCCT GCTGGGATTA TATCAACAAC CAGTGTGATG AGCTGTGCAA CACGGTCGAG TGCCTGTTTG ACAACTTTGA ATGCCAGGGG
 P  L  P    C  W  D  Y    I  N  N    Q  C  D    E  L  C  N    T  V  E    C  L  F    D  N  F  E    C  Q  G>

           4510        4520        4530        4540        4550        4560        4570        4580        4590
            •           •           •           •           •           •           •           •           •
AACAGCAAGA CATGCAAGTA TGACAAATAC TGTGCAGACC ACTTCAAAGA CAACCACTGT AACCAGGGGT GCAACAGTGA GGAGTGTGGT
 N  S  K    T  C  K  Y    D  K  Y    C  A  D    H  F  K  D    N  H  C    N  Q  G    C  N  S  E    E  C  G>

           4600        4610        4620        4630        4640        4650        4660        4670        4680
            •           •           •           •           •           •           •           •           •
TGGGATGGGC TGGACTGTGC TGCTGACCAA CCTGAGAACC TGGCAGAAGG TACCCTGGTT ATTGTGGTAT TGATGCCACC TGAACAACTG
 W  D  G    L  D  C  A    A  D  Q    P  E  N    L  A  E  G    T  L  V    I  V  V    L  M  P  P    E  Q  L>

           4690        4700        4710        4720        4730        4740        4750        4760        4770
            •           •           •           •           •           •           •           •           •
CTCCAGGATG CTCGCAGCTT CTTGCGGGCA CTGGGTACCC TGCTCCACAC CAACCTGCGC ATTAAGCGGG ACTCCCAGGG GGAACTCATG
 L  Q  D    A  R  S  F    L  R  A    L  G  T    L  L  H  T    N  L  R    I  K  R    D  S  Q  G    E  L  M>

           4780        4790        4800        4810        4820        4830        4840        4850        4860
            •           •           •           •           •           •           •           •           •
GTGTACCCCT ATTATGGTGA GAAGTCAGCT GCTATGAAGA AACAGAGGAT GACACGCAGA TCCCTTCCTG GTGAACAAGA ACAGGAGGTG
 V  Y  P    Y  Y  G  E    K  S  A    A  M  K    K  Q  R  M    T  R  R    S  L  P    G  E  Q  E    Q  E  V>
```

<p align="center">FIG.17F</p>

```
        4870       4880       4890       4900       4910       4920       4930       4940       4950
         •          •          •          •          •          •          •          •          •
GCTGGCTCTA AAGTCTTTCT GGAAATTGAC AACCGCCAGT GTGTTCAAGA CTCAGACCAC TGCTTCAAGA ACACGGATGC AGCAGCAGCT
 A  G  S   K  V  F  L   E  I  D   N  R  Q   C  V  Q  D   S  D  H   C  F  K   N  T  D  A   A  A  A>

        4960       4970       4980       4990       5000       5010       5020       5030       5040
         •          •          •          •          •          •          •          •          •
CTCCTGGCCT CTCACGCCAT ACAGGGGACC CTGTCATACC CTCTTGTGTC TGTCGTCAGT GAATCCCTGA CTCCAGAACG CACTCAGCTC
 L  L  A   S  H  A  I   Q  G  T   L  S  Y   P  L  V  S   V  V  S   E  S  L   T  P  E  R   T  Q  L>

        5050       5060       5070       5080       5090       5100       5110       5120       5130
         •          •          •          •          •          •          •          •          •
CTCTATCTCC TTGCTGTTGC TGTTGTCATC ATTCTGTTTA TTATTCTGCT GGGGGTAATC ATGGCAAAAC GAAAGCGTAA GCATGGCTCT
 L  Y  L   L  A  V  A   V  V  I   I  L  F   I  I  L  L   G  V  I   M  A  K  R   K  R  K   H  G  S>

        5140       5150       5160       5170       5180       5190       5200       5210       5220
         •          •          •          •          •          •          •          •          •
CTCTGGCTGC CTGAAGGTTT CACTCTTCGC CGAGATGCAA GCAATCACAA GCGTCGTGAG CCAGTGGGAC AGGATGCTGT GGGGCTGAAA
 L  W  L   P  E  G  F   T  L  R   R  D  A   S  N  H  K   R  R  E   P  V  G   Q  D  A  V   G  L  K>

        5230       5240       5250       5260       5270       5280       5290       5300       5310
         •          •          •          •          •          •          •          •          •
AATCTCTCAG TGCAAGTCTC AGAAGCTAAC CTAATTGGTA CTGGAACAAG TGAACACTGG GTCGATGATG AAGGGCCCCA GCCAAAGAAA
 N  L  S   V  Q  V  S   E  A  N   L  I  G   T  G  T  S   E  H  W   V  D  D   E  G  P  Q   P  K  K>

        5320       5330       5340       5350       5360       5370       5380       5390       5400
         •          •          •          •          •          •          •          •          •
GTAAAGGCTG AAGATGAGGC CTTACTCTCA GAAGAAGATG ACCCCATTGA TCGACGGCCA TGGACACAGC AGCACCTTGA AGCTGCAGAC
 V  K  A   E  D  E  A   L  L  S   E  E  D   D  P  I  D   R  R  P   W  T  Q   Q  H  L  E   A  A  D>

        5410       5420       5430       5440       5450       5460       5470       5480       5490
         •          •          •          •          •          •          •          •          •
ATCCGTAGGA CACCATCGCT GGCTCTCACC CCTCCTCAGG CAGAGCAGGA GGTGGATGTG TTAGATGTGA ATGTCCGTGG CCCAGATGGC
 I  R  R   T  P  S  L   A  L  T   P  P  Q   A  E  Q  E   V  D  V   L  D  V   N  V  R  G   P  D  G>

        5500       5510       5520       5530       5540       5550       5560       5570       5580
         •          •          •          •          •          •          •          •          •
TGCACCCCAT TGATGTTGGC TTCTCTCCGA GGAGGCAGCT CAGATTTGAG TGATGAAGAT GAAGATGCAG AGGACTCTTC TGCTAACATC
 C  T  P   L  M  L  A   S  L  R   G  G  S   S  D  L  S   D  E  D   E  D  A   E  D  S  S   A  N  I>

        5590       5600       5610       5620       5630       5640       5650       5660       5670
         •          •          •          •          •          •          •          •          •
ATCACAGACT TGGTCTACCA GGGTGCCAGC CTCCAGGCCC AGACAGACCG GACTGGTGAG ATGGCCCTGC ACCTTGCAGC CCGCTACTCA
 I  T  D   L  V  Y  Q   G  A  S   L  Q  A   Q  T  D  R   T  G  E   M  A  L   H  L  A  A   R  Y  S>
```

FIG.17G

63

```
        5680      5690      5700      5710      5720      5730      5740      5750      5760
CGGGCTGATC CTGCCAAGCG TCTCCTGGAT GCAGGTGCAG ATGCCAATGC CCAGGACAAC ATGGGCCGCT GTCCACTCCA TGCTGCAGTG
  R  A  D   A  A  K  R   L  L  D   A  G  A   D  A  N  A   Q  D  N   M  G  R   C  P  L  H   A  A  V>

        5770      5780      5790      5800      5810      5820      5830      5840      5850
GCAGCTGATG CCCAAGGTGT CTTCCAGATT CTGATTCGCA ACCGAGTAAC TGATCTAGAT GCCAGGATGA ATGATGGTAC TACACCCCTG
  A  A  D   A  Q  G  V   F  Q  I   L  I  R   N  R  V  T   D  L  D   A  R  M   N  D  G  T   T  P  L>

        5860      5870      5880      5890      5900      5910      5920      5930      5940
ATCCTGGCTG CCCGCCTGGC TGTGGAGGGA ATGGTGGCAG AACTGATCAA CTGCCAAGCG GATGTGAATG CAGTGGATGA CCATGGAAAA
  I  L  A   A  R  L  A   V  E  G   M  V  A   E  L  I  N   C  Q  A   D  V  N   A  V  D  D   H  G  K>

        5950      5960      5970      5980      5990      6000      6010      6020      6030
TCTGCTCTTC ACTGGGCAGC TGCTGTCAAT AATGTGGAGG CAACTCTTTT GTTGTTGAAA AATGGGGCCA ACCGAGACAT GCAGGACAAC
  S  A  L   H  W  A  A   A  V  N   N  V  E   A  T  L  L   L  L  K   N  G  A   N  R  D  M   Q  D  N>

        6040      6050      6060      6070      6080      6090      6100      6110      6120
AAGGAAGAGA CACCTCTGTT TCTTGCTGCC CGGGAGGGGA GCTATGAAGC AGCCAAGATC CTGTTAGACC ATTTTGCCAA TCGAGACATC
  K  E  E   T  P  L  F   L  A  A   R  E  G   S  Y  E  A   A  K  I   L  L  D   H  F  A  N   R  D  I>

        6130      6140      6150      6160      6170      6180      6190      6200      6210
ACAGACCATA TGGATCGTCT TCCCCGGGAT GTGGCTCGGG ATCGCATGCA CCATGACATT GTGCGCCTTC TGGATGAATA CAATGTGACC
  T  D  H   M  D  R  L   P  R  D   V  A  R   D  R  M  H   H  D  I   V  R  L   L  D  E  Y   N  V  T>

        6220      6230      6240      6250      6260      6270      6280      6290      6300
CCAAGCCCTC CAGGCACCGT GTTGACTTCT GCTCTCTCAC CTGTCATCTG TGGGCCCAAC AGATCTTTCC TCAGCCTGAA GCACACCCCA
  P  S  P   P  G  T  V   L  T  S   A  L  S   P  V  I  C   G  P  N   R  S  F   L  S  L  K   H  T  P>

        6310      6320      6340      6350      6360      6370      6380      6390      6400
ATGGGCAAGA AGTCTAGACG GCCCAGTGCC AAGAGTACCA TGCCTACTAG CCTCCCTAAC CTTGCCAAGG AGGCAAAGGA TGCCAAGGGT
  M  G  K   K  S  R  R   P  S  A   K  S  T   M  P  T  S   L  P  N   L  A  K   E  A  K  D   A  K  G>

        6400      6410      6420      6430      6440      6450      6460      6470      6480
AGTAGGAGGA AGAAGTCTCT GAGTGAGAAG GTCCAACTGT CTGAGAGTTC AGTAACTTTA TCCCCTGTTG ATTCCCTAGA ATCTCCTCAC
  S  R  R   K  K  S  L   S  E  K   V  Q  L   S  E  S  S   V  T  L   S  P  V   D  S  L  E   S  P  H>
```

FIG.17H

```
        6490        6500        6510        6520        6530        6540        6550        6560        6570
          •           •           •           •           •           •           •           •           •
ACGTATGTTT CCGACACCAC ATCCTCTCCA ATGATTACAT CCCCTGGGAT CTTACAGGCC TCACCCAACC CTATGTTGGC CACTGCCCGC
  T  Y  V    S  D  T  T    S  S  P    M  I  T    S  P  G  I    L  Q  A    S  P  N    P  M  L  A    T  A  A>

        6580        6590        6600        6610        6620        6630        6640        6650        6660
          •           •           •           •           •           •           •           •           •
CCTCCTGCCC CAGTCCATGC CCAGCATGCA CTATCTTTTT CTAACCTTCA TGAAATGCAG CCTTTGGCAC ATGGGGCCAG CACTGTGCTT
  P  P  A    P  V  H  A    Q  H  A    L  S  F    S  N  L    H  E  M  Q    P  L  A    H  G  A  S    T  V  L>

        6670        6680        6690        6700        6710        6720        6730        6740        6750
          •           •           •           •           •           •           •           •           •
CCCTCAGTGA GCCAGTTGCT ATCCCACCAC CACATTGTGT CTCCAGGCAG TGGCAGTGCT GGAAGCTTGA GTAGGCTCCA TCCAGTCCCA
  P  S  V    S  Q  L  L    S  H  H    H  I  V    S  P  G  S    G  S  A    G  S  L    S  R  L  H    P  V  P>

        6760        6770        6780        6790        6800        6810        6820        6830        6840
          •           •           •           •           •           •           •           •           •
GTCCCAGCAG ATTGGATGAA CCGCATGGAG GTGAATGAGA CCCAGTACAA TGAGATGTTT GGTATGGTCC TGGCTCCAGC TGAGGGCACC
  V  P  A    D  W  M  N    R  M  E    V  N  E    T  Q  Y  N    E  M  F    G  M  V    L  A  P  A    E  G  T>

        6850        6860        6870        6880        6890        6900        6910        6920        6930
          •           •           •           •           •           •           •           •           •
CATCCTGGCA TAGCTCCCCA GAGCAGGCCA CCTGAAGGGA AGCACATAAC CACCCCTCGG GAGCCCTTGC CCCCCATTGT GACTTTCCAG
  H  P  G    I  A  P  Q    S  R  P    P  E  G    K  H  I  T    T  P  R    E  P  L    P  P  I  V    T  F  Q>

        6940        6950        6960        6970        6980        6990        7000        7010        7020
          •           •           •           •           •           •           •           •           •
CTCATCCCTA AAGGCAGTAT TGCCCAACCA GCGGGGGCTC CCCAGCCTCA GTCCACCTGC CCTCCAGCTG TTGCGGGGCCC CCTGCCCACC
  L  I  P    K  G  S  I    A  Q  P    A  G  A    P  Q  P  Q    S  T  C    P  P  A    V  G  P    L  P  T>

        7030        7040        7050        7060        7070        7080        7090        7100        7110
          •           •           •           •           •           •           •           •           •
ATGTACCAGA TTCCAGAAAT GGCCCGTTTG CCCAGTGTGG CTTTCCCCAC TGCCATGATG CCCCAGCAGG ACGGGCAGGT AGCTCAGACC
  M  Y  Q    I  P  E  M    A  R  L    P  S  V    A  F  P  T    A  M  M    P  Q  Q    D  G  Q  V    A  Q  T>

        7120        7130        7140        7150        7160        7170        7180        7190        7200
          •           •           •           •           •           •           •           •           •
ATTCTCCCAG CCTATCATCC TTTCCCAGCC TCTGTGGGCA AGTACCCCAC ACCCCCTTCA CAGCACAGTT ATGCTTCCTC AAATGCTGCT
  I  L  P    A  Y  H  P    F  P  A    S  V  G    K  Y  P  T    P  P  S    Q  H  S    Y  A  S  S    N  A  A>

        7210        7220        7230        7240        7250        7260        7270        7280        7290
          •           •           •           •           •           •           •           •           •
GAGCGAACAC CCAGTCACAG TGGTCACCTC CAGGGTGAGC ATCCCTACCT GACACCATCC CCAGAGTCTC CTGACCAGTG GTCAAGTTCA
  E  R  T    P  S  H  S    G  H  L    Q  G  E    H  P  Y  L    T  P  S    P  E  S    P  D  Q  W    S  S  S>
```

FIG.171

```
        7300       7310       7320       7330       7340       7350       7360       7370       7380
         •          •          •          •          •          •          •          •          •
TCACCCCACT CTGCTTCTGA CTGGTCAGAT GTGACCACCA GCCCTACCCC TGGGGGTGCT GGAGGAGGTC AGCGGGGACC TGGGACACAC
 S  P  H    S  A  S  D    W  S  D    V  T  T    S  P  T  P    G  G  A    G  G  G    Q  R  G  P    G  T  H>

        7390       7400       7410       7420       7430       7440       7450       7460       7470
         •          •          •          •          •          •          •          •          •
ATGTCTGAGC CACCACACAA CAACATGCAG GTTTATGCGT GAGAGAGTCC ACCTCCAGTG TAGAGACATA ACTGACTTTT GTAAATGCTG
 M  S  E    P  P  H  N    N  M  Q    V  Y  A>

        7480       7490       7500       7510       7520       7530       7540       7550       7560
         •          •          •          •          •          •          •          •          •
CTGAGGAACA AATGAAGGTC ATCCGGGAGA GAAATGAAGA AATCTCTGGA GCCAGCTTCT AGAGGTAGGA AAGAGAAGAT GTTCTTATTC

        7570       7580       7590       7600       7610       7620       7630       7640       7650
         •          •          •          •          •          •          •          •          •
AGATAATGCA AGAGAAGCAA TTCGTCAGTT TCACTGGGTA TCTGCAAGGC TTATTGATTA TTCTAATCTA ATAAGACAAG TTTGTGGAAA

        7660       7670       7680       7690       7700       7710       7720       7730       7740
         •          •          •          •          •          •          •          •          •
TGCAAGATGA ATACAAGCCT TGGGTCCATG TTTACTCTCT TCTATTTGGA GAATAAGATG GATGCTTATT GAAGCCCAGA CATTCTTGCA

        7750       7760       7770       7780       7790       7800       7810       7820       7830
         •          •          •          •          •          •          •          •          •
GCTTGGACTG CATTTTAAGC CCTGCAGGCT TCTGCCATAT CCATGAGAAG ATTCTACACT AGCGTCCTGT TGGGAATTAT GCCCTGGAAT

        7840       7850       7860       7870       7880       7890       7900       7910       7920
         •          •          •          •          •          •          •          •          •
TCTGCCTGAA TTGACCTACG CATCTCCTCC TCCTTGGACA TTCTTTTGTC TTCATTTGGT GCTTTTGGTT TTGCACCTCT CCGTGATTGT

        7930       7940       7950       7960       7970       7980       7990       8000       8010
         •          •          •          •          •          •          •          •          •
AGCCCTACCA GCATGTTATA GGGCAAGACC TTTGTGCTTT TGATCATTCT GGCCCATGAA AGCAACTTTG GTCTCCTTTC CCCTCCTGTC

        8020       8030       8040       8050       8060       8070       8080       8090       8100
         •          •          •          •          •          •          •          •          •
TTCCCGGTAT CCCTTGGAGT CTCACAAGGT TTACTTTGGT ATGGTTCTCA GCACAAACCT TTCAAGTATG TTGTTTCTTT GGAAAATGGA

        8110       8120       8130       8140       8150       8160       8170       8180       8190
         •          •          •          •          •          •          •          •          •
CATACTGTAT TGTGTTCTCC TGCATATATC ATTCCTGGAG AGAGAAGGGG AGAAGAATAC TTTTCTTCAA CAAATTTTGG GGGCAGGAGA

        8200       8210       8220       8230       8240       8250       8260       8270       8280
         •          •          •          •          •          •          •          •          •
TCCCTTCAAG AGGCTGCACC TTAATTTTTC TTGTCTGTGT GCAGGTCTTC ATATAAACTT TACCAGGAAG AAGGGTGTGA GTTTGTTGTT
```

FIG.17J

```
        8290        8300        8310        8320        8330        8340        8350        8360        8370
          •           •           •           •           •           •           •           •           •
TTTCTGTGTA TGGGCCTGGT CAGTGTAAAG TTTTATCCTT GATAGTCTAG TTACTATGAC CCTCCCCACT TTTTTAAAAC CAGAAAAAGG


        8380        8390        8400        8410        8420        8430        8440        8450        8460
          •           •           •           •           •           •           •           •           •
TTTGGAATGT TGGAATGACC AAGAGACAAG TTAACTCGTG CAAGAGCCAG TTACCCACCC ACAGGTCCCC CTACTTCCTG CCAAGCATTC


        8470        8480        8490        8500        8510        8520        8530        8540        8550
          •           •           •           •           •           •           •           •           •
CATTGACTGC CTGTATGGAA CACATTTGTC CCAGATCTGA GCATTCTAGG CCTGTTTCAC TCACTCACCC AGCATATGAA ACTAGTCTTA


        8560        8570        8580        8590        8600        8610        8620        8630        8640
          •           •           •           •           •           •           •           •           •
ACTGTTGAGC CTTTCCTTTC ATATCCACAG AAGACACTGT CTCAAATGTT GTACCCTTGC CATTTAGGAC TGAACTTTCC TTAGCCCAAG


        8650        8660        8670        8680        8690        8700        8710        8720        8730
          •           •           •           •           •           •           •           •           •
GGACCCAGTG ACAGTTGTCT TCCGTTTGTC AGATGATCAG TCTCTACTGA TTATCTTGCT GCTTAAAGGC CTGCTCACCA ATCTTTCTTT


        8740        8750        8760        8770        8780        8790        8800        8810        8820
          •           •           •           •           •           •           •           •           •
CACACCGTGT GGTCCGTGTT ACTGGTATAC CCAGTATGTT CTCACTGAAG ACATGGACTT TATATGTTCA AGTGCAGGAA TTGGAAAGTT


        8830        8840        8850        8860        8870        8880        8890        8900        8910
          •           •           •           •           •           •           •           •           •
GGACTTGTTT TCTATGATCC AAAACAGCCC TATAAGAAGG TTGGAAAAGG AGGAACTATA TAGCAGCCTT TGCTATTTTC TGCTACCATT


        8920        8930        8940        8950        8960        8970        8980        8990        9000
          •           •           •           •           •           •           •           •           •
TCTTTTCCTC TGAAGCGGCC ATGACATTCC CTTTGGCAAC TAACGTAGAA ACTCAACAGA ACATTTTCCT TTCCTAGAGT CACCTTTTAG


        9010        9020        9030        9040        9050        9060        9070        9080        9090
          •           •           •           •           •           •           •           •           •
ATGATAATGG ACAACTATAG ACTTGCTCAT TGTTCAGACT GATTGCCCCT CACCTGAATC CACTCTCTGT ATTCATGCTC TTGGCAATTT


        9100        9110        9120        9130        9140        9150        9160        9170        9180
          •           •           •           •           •           •           •           •           •
CTTTGACTTT CTTTTAAGGG CAGAAGCATT TTAGTTAATT GTAGATAAAG AATAGTTTTC TTCCTCTTCT CCTTGGGCCA GTTAATAATT


        9190        9200        9210        9220        9230        9240        9250        9260        9270
          •           •           •           •           •           •           •           •           •
GGTCCATGGC TACACTGCAA CTTCCGTCCA GTGCTGTGAT GCCCATGACA CCTGCAAAAT AAGTTCTGCC TGGGCATTTT GTAGATATTA
```

FIG.17K

```
        9280        9290        9300        9310        9320        9330        9340        9350        9360
         •           •           •           •           •           •           •           •           •
 ACAGGTGAAT TCCCGACTCT TTTGGTTTGA ATGACAGTTC TCATTCCTTC TATGGCTGCA AGTATGCATC AGTGCTTCCC ACTTACCTGA


        9370        9380        9390        9400        9410        9420        9430        9440        9450
         •           •           •           •           •           •           •           •           •
 TTTGTCTGTC GGTGGCCCCA TATGGAAACC CTGCCGTGTCT GTTGGCATAA TAGTTTACAA ATGGTTTTTT CAGTCCTATC CAAATTTATT


        9460        9470        9480        9490        9500        9510        9520        9530        9540
         •           •           •           •           •           •           •           •           •
 GAACCAACAA AAATAATTAC TTCTGCCCTG AGATAAGCAG ATTAAGTTTG TTCATTCTCT GCTTTATTCT CTCCATGTGG CAACATTCTG


        9550        9560        9570        9580        9590        9600        9610        9620        9630
         •           •           •           •           •           •           •           •           •
 TCAGCCTCTT TCATAGTGTG CAAACATTTT ATCATTCTAA ATGGTGACTC TCTGCCCTTG GACCCATTTA TTATTCACAG ATGGGGAGAA


        9640        9650        9660        9670        9680        9690        9700        9710        9720
         •           •           •           •           •           •           •           •           •
 CCTATCTGCA TGGACCCTCA CCATCCTCTG TGCAGCACAC ACAGTGCAGG GAGCCAGTGG CGATGGCGAT GACTTTCTTC CCCTGGGAAT

 TCC
```

FIG.17L

68